PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/00, 15/12, 1/21, C07K 13/00, 15/28, C12Q 1/68, A01K 67/027 — A1

(11) International Publication Number: **WO 94/13794**

(43) International Publication Date: 23 June 1994 (23.06.94)

(21) International Application Number: PCT/US93/11725

(22) International Filing Date: 3 December 1993 (03.12.93)

(30) Priority Data:
07/985,296 4 December 1992 (04.12.92) US

(71) Applicant: BRIGHAM AND WOMEN'S HOSPITAL, INC. [US/US]; 75 Francis Street, Boston, MA 02115 (US).

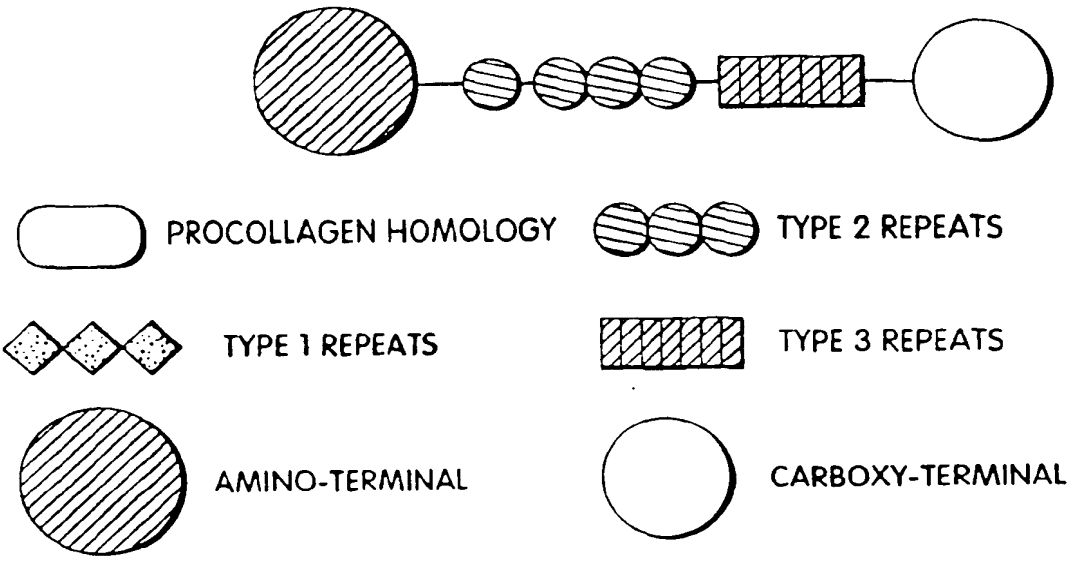
(72) Inventor: LAWLER, John, W.; 657 Humphrey Street, Swampscott, MA 01907 (US).

(74) Agent: GATES, Edward, R.; Wolf, Greenfield & Sacks, 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: HUMAN THROMBOSPONDIN-4



(57) Abstract

A novel member of the thrombospondin gene family, thrombospondin-4, has been cloned and sequenced. A frog thrombospondin-4 DNA and the mammalian homolog of the frog DNA are disclosed. Recombinant vectors and cells are described. Methods of providing isolated thrombospondin-4 DNA and polypeptide sequences are disclosed, as well as methods of making transgenic animals containing, or lacking, the thrombospondin-4 gene.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

HUMAN THROMBOSPONDIN-4

This invention made with U.S. Government Support under National Institutes of Health Grant No. NIH: HL28749. The U.S. Government has certain rights to this invention.

BACKGROUND OF THE INVENTION

Platelet thrombospondin is a glycoprotein that is structurally and functionally similar to the adhesive glycoproteins found in a wide variety of cells. The thrombospondin genes encode two distinct polypeptides, designated thrombospondin -1 and -2 (Bornstein et al., J. Biol. Chem., 266:12821-12824, 1991; and 265:16691-16698, 1991; Proc. Nat. Acad. Sci. USA 88:8636-8640 (1990); Wolf et al., Genomics, 6:685-691 1990)). Thrombospondin-3 is a recently discovered member of the thrombospondin gene family (Vos et al. J. Biol Chem, 267: 12192-12196 (1992)).

Partial or complete cDNA sequences are available for human, mouse and frog thrombospondin-1, and human, mouse and chicken thrombospondin-2 (Lawler and Hynes, J. Cell Biol., 103:1635-1648; (1986); Bornstein et al., supra; Lawler et al., J. Biol. Chem., 266:8039-8043 (1991); Genomics, 11: 587-600, (1991). The overall molecular architecture of thrombospondin-1 and 2 are substantially the same. The predicted amino acid sequences of thrombospondins-1 and -2 are very similar in their repeat sequences and their COOH-terminal domains.

The central portion of platelet thrombospondin is composed of mutiple copies of structural motifs found in other proteins (Lawler and Hynes, supra 1986). Amino acid sequences that have been shown to mediate cellular attachment are also present in the central portion of the molecule (Rich et al., Science, 249. 1574-1577 (1990)). In addition, thrombospondin contains a region that is rich in calcium

binding sites and that contains the RGD sequence that promotes adhesion of some cell types (Lawler et al., (1988)).

Thrombospondin has been shown to modulate its attachment to a variety of cell types in vitro. The NH2-terminal heparin-binding domain binds to proteoglycans including syndecan and to cell surface sulfatides; (Sun et al., J. Biol. Chem., 264:2885-1889 (1989)). Thrombospondin also interacts with CD36 or platelet glycoprotein IV (Stromski et al., Exp. Cell Res., 198:85-92 (1992)). Several integrin receptors have been reported to bind thrombospondin (Lawler et al., supra (1988)). These integrin receptors are reported to be involved in neurite outgrowth (Neugebauer, et al., Neuron, b:345-358 (1991)). Through these, and yet to be identified interactions, thrombospondin can modulate cell adhesion, cell migration, angiogenesis and neurite outgrowth.

The human platelet thrombospondins 1 and 2 that have already been characterized in the prior art are schematically illustrated in FIG. 1. The term "thrombospondin" refers to adhesive glycoproteins of about 420,000-dalton molecular weight that are involved in modulation of cell growth and migration. Thrombospondins are composed of three polypeptides linked by disulfide bonds. The N-terminal end binds with heparin, the C-terminal end assists in platelet aggregation.

Three types of internal repeating structures are found in human thrombospondin-1 and thrombospondin-2 polypeptides. These are the type 1, 2 and 3 domains ("repeats"). In addition to the three types of domains, thrombospondins 1 and 2 also contain a region of homology to procollagen, as well as amino and carboxyl-termini.

Human thrombospondins-1 and -2 have three, type 1 domains. Type 1 domains are homologous to several of the complement factors, including C-8, C-9 and properdin. Type 1 domains are also found in two proteins produced in malaria-parasitized blood cells. These are circumsporozoite protein and the thrombospondin related anonymous protein

(Robson et al., Nature 335: 79-82, (1988)). Three copies of type 1 domains are also found in the UNC-5 gene of C. elegans (Culotti, et al. J. Cell Biol. 115: 1229, (1991)). The type 1 domains of thrombospondin-1 and -2 extend from nucleic acid number 1210 to 1719 (Lawler and Hynes, J. Cell Biol. 103: 1635 (1986)).

Human thrombospondins-1 and 2 have three, type 2 domains. Type 2 domains are similar to epidermal growth factor (EGF) in that they are framed around a characteristic spacing of six cysteines. Multiple copies of EGF repeat are commonly found in adhesive glycoproteins and cell adhesion molecules. The type 2 domains extend from nucleic acid sequence 1720to 2151 on thrombospondins-1 and -2.

SUMMARY OF THE INVENTION

According to one aspect of the invention, an isolated nucleotide sequence encoding a new member of the thrombospondin family, thrombospondin-4, or unique fragments of thrombospondin-4, is provided. One embodiment is an isolated DNA sequence encoding a thrombospondin, that has at least four, type 2 domains. In another embodiment, the sequence encodes a thrombospondin that lacks any type 1 domains. A further embodiment is a sequence encoding a thrombospondin that lacks a region of homology with procollagen. Yet another embodiment is a sequence that encodes a thrombospondin that has four, type 2 domains, lacks type 1 domains and lacks a region of homology to procollagen. The preferred DNA of the present invention is a human homolog of thrombospondin-4. Additionally, the invention relates to vertebrate thrombospondin-4 genes isolated from porcine, ovine, bovine, feline, avian, equine, or canine, as well as primate sources and any other species in which thrombospondin-4 structure exists.

Also provided are recombinant cells and plasmids containing the foregoing isolated DNA, preferably linked to a promoter. Portions of the foregoing nucleotide sequences are

also included in the invention. One such portion is contained in a vector within a host cell.

According to another aspect of the invention, isolated thrombospondin protein is provided, having at least four type 2 domains. Other thrombospondins lack any type 1 domains and/or lack any procollagen homology. Portions of the foregoing isolated thrombospondin proteins are also included in the invention. Antibodies with selective binding specificity for the thrombospondin protein of the invention also are provided.

Another aspect of the invention is a method for producing thrombospondin polypeptide. The method includes providing an expression vector to a host, the vector containing a DNA sequence of the invention having at least four, type 2 domains; allowing the host to express the thrombospondin, and isolating the expressed thrombospondin.

A further aspect of the invention is a probe capable of distinguishing thrombospondin-4 from thrombospondins -1, -2, and -3. The probe can include a nucleotide sequence encoding a thrombospondin-4 polypeptide with at least four, type 2 domains, that lacks any type 1 domains, and lacks a region of homology to procollagen. The nucleotide sequence also can encode a thrombospondin-4 polypeptide having sequences unique to the polypeptide.

Also provided is a thrombospondin-4 polypeptide having a restricted range of expression in tissues. The preferred polypeptide is expressed in human heart and skeletal muscle, but is not expressed in human placenta, liver or kidney.

The novel molecules of the invention can be employed in experimental or therapeutic protocols. For example, a method for interfering with the activity of a thrombospondin-4 gene may be accomplished by providing a construct arranged to include a thrombospondin nucleotide sequence which, when inserted, inactivates either transcription of messenger for thrombospondin-4 and/or inactivates translation of messenger into thrombospondin-4 protein. This construct further has a

promotor operatively linked to the sequence. Next, the construct is introduced into a cell, and the construct is allowed to homologously recombine with complementary sequences of the cell genome. Finally, cells lacking the ability to transcribe thrombospondin-4 are selected.

These and other aspects of the invention as well as various advantages in the utilities will be more apparent with reference to the detailed description of the invention when taken in connection with the accompanying drawings. It is to be understood that the drawings are designed for the purpose of illustration only and are not intended as a definition of the limits of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1: Schematic drawing of human thrombospondin-1 and thrombospondin-2.

FIG. 2: Schematic drawing of human thrombospondin-4 The drawing schematically depicts an actual nucleotide sequence of 3120 nucleotides, with a message of approximately 3.3 kb.

FIG. 3: Alignment of restriction fragments of Xenopus thrombospondin-4 clones. Restriction endonuclease sites are indicated for the two families (TSP-4A and TSP-4B). The clones that have been isolated in the first (XF1-XF4), second (XS5-XS10) and third (XT11-XT14) rounds of screening have been grouped into their appropriate family by restriction endonuclease mapping and nucleotide sequencing.

Fig. 4: Photograph of a Northern blot of Xenopus stage 17 RNA probed with the XF3 clone of Fig. 3. Two micrograms of total stage 17 mRNA was electrophoresed and blotted. Positions and sizes of markers are shown on the left.

FIG. 5: The expression of thrombospondin-4 in adult human tissue. A northern blot of poly A + RNA from adult human heart (a), brain (b), placenta (c), lung (d), liver (e), skeletal muscle (f), kidney (g) and pancreas (h). The blot was probed with a 2.2 kb fragment of Xenopus thrombospondin-4. The positions and sizes (kb) of the markers are indicated on the left.

DETAILED DESCRIPTION OF THE INVENTION

Human thrombospondins 1 and 2 have seven, type 3 domains. Type 3 domains extend from nucleic acid 2221 to 2926 on thrombospondins -1 and -2. Type 3 domains include a large number of calcium-binding sites. The consensus sequence of these type 3 domains is similar to calcium binding site sequences of calmodulin, parvalbumin and fibrinogen beta and gamma subunits. (Lawler and Hynes, supra). In particular, there are aspartic acid residues at positions 6, 8, 10, 14 and 17 of the type 3 domains, as well as a second set at positions 21, 23, 25, 29 and 32. Moreover, glycine residues at positions 11 and 26 are also homologous with calcium-binding sites of calmodulin and paravalbumin. Lawler and Hynes, supra. To date, no other protein has been identified that could potentially bind as much calcium as thrombospondin. Furthermore, no other protein has been identified in which the calcium binding sites are contiguous. The thrombospondins of the invention, like other thrombospondins characterize to date (i.e. thrombospondins -1 -2), have an N-terminal region that is more than 200 amino acids in length. In thrombospondins -3 and -4, which lack procollagen and type 1 domains, this N-terminal region preceeds the type 2 domains. In thrombospondins -1 and -2, this N-terminal region preceeds both the procollagen and type 1 domains.

Thrombospondins -1 and -2 also have a region adjacent the N-terminal end that is substantially homologous to the known sequence of procollagen. This region extends from nucleotides 916 to 1209 on thrombospondins -1 and -2.

The novel member of the thrombospondin family, hereinafter called "thrombospondin-4" has the schematic structure depicted in FIG. 2.

In complete contrast to human thrombospondins 1 and 2, thrombospondin-4 lacks type 1 domains entirely. Thrombospondin-4 also lacks a region homologous to procollagen, in contrast to the known thrombospondins 1 and

2. The molecular architecture of much of the N-terminal end of thrombospondin-4 is thus distinct from that of human thrombospondins 1 or 2.

Moreover, thrombospondin-4 has four, type 2 domains (FIG. 2) in contrast to thrombospondins -1 and -2 which have three, type 2 domains (see FIG. 1).

Thrombospondin-4 has the same number of calcium-binding sites located within the type 3 domains as do thrombospondins 1 and 2.

The configuration and number of repeats, as well as the lack of procollagen homology and lack of type 1 domains, define the unique thrombospondin-4 structure.

One embodiment of a thrombospondin-4 molecule, according to the invention, is the isolated nucleotide sequence shown in SEQ ID NO.: 1. By "isolated" it is meant a nucleic acid sequence: (i) amplified in vitro by, for example, polymerase chain reaction (PCR); (ii) synthesized by, for example, chemical synthesis; (iii) recombinantly produced by cloning; or (iv) purified, as by cleavage and gel separation. The term "isolated" is also meant to include polypeptides encoded by isolated nucleic acid sequences, as well as polypeptides synthesized by, for example, chemical synthetic methods, and polypeptides separated from biological materials, and then purified using conventional protein analytical procedures.

SEQ ID NO.: 1 is a thrombospondin-4 that has been isolated from the frog, Xenopus laevis.
An open reading frame of 889 amino acids is predicted from the Xenopus nucleotide sequence. The deduced amino acid sequence encoded by the Xenopus thrombospondin-4 DNA sequence is given in SEQ ID NO.: 2. The first 216 amino acids of Xenopus thrombospondin-4 have little homology with human thrombospondins 1 and 2, primarily because of the lack of type 1 repeats and the lack of procollagen sequence in Xenopus thrombospondin-4.

Four adjacent type 2 domains can be identified in Xenopus thrombospondin-4 on the basis of the positions of the

cysteine residues. The overall homology with other thrombospondins is low in this type 2 region, and the introduction of several gaps is necessary to optimize the alignment. The second of the type 2 domains is, however, similar to those of thrombospondins -1 and -2, in that thirteen residues are inserted between the last two cysteine residues. The amino acid sequence for the four type 2 domains of thrombospondin-4 are shown below in Table 1.

Table 1: TYPE 2 DOMAINS OF THROMBOSPONDIN-4

PRCDATS---CFRGVRCIDTEGGFQ-CGPCPEGYTG----------NGVICTDV
DECRL--NP-CFLGVRCINTSPGFK-CESCPPGYTGSTIQGIGINFAKQNKQVCTDT
NECENGRNGGCTSNSLCINTMGSFR-CGGCKPGYVG----------DQIKGCKPE
KSCRHGQNP-CHASAQCSEEKVGDVTCT-CSVGWAG----------NGYLCGK

The type 3 domains of Xenopus thrombospondin-4 are 61.4% identical to the type 3 domains of human thrombospondins 1 and 2. The consensus sequence and overall organization of the seven, type 3 repeats of Xenopus thrombospondin-4 are equivalent to those of thrombospondins-1 and -2, with the second and fourth type 3 domains being truncated after the second cysteine. Thrombospondin-4, however, contains 4 amino acids (PPGP) at the end of the sixth, type 3 domain that do not align with sequences on thrombospondins -1 and -2. Further thrombospondin-4 does not contain an RGD sequence. The seven, type 3 domains of Xenopus thrombospondin-4 are shown below in Table 2.

The consensus sequence for Xenopus is compared to that for human and mouse thrombospondin -1 and chicken thrombospondin -2 at the bottom of Table 2. The underline indicates that an N occupies one of the positions that is occupied by a D.

Table 2: TYPE 3 DOMAINS OF THROMBOSPONDIN-4

DNCVYVPNSGQEDTDKDNIGDACDE--DADGDGILNEQ	
DNCVLAANIDQKNSDQDIFGDAC	
DNCRLTLNNDQRDTDNDGKGDACDD--DMDGDGIKNIL	
DNCQRVPNVDQKDKDGDGVGDIC	
DSCPDIINPNQSDIDNDLVGDSCDTNQDSDGDGHQDST	
DNCPTVINSNQLDTDKDGIGDECDD--DDDNDGIPDTVPPGP	
DNCKLVPNPGQEDDNNDGVGDVCEA--DFDQDTVIDRI	
D.C....N..Q.D.D.D..GD.C....D.D.D.....	Consensus
DNC....N..Q.D.D.D..GD.C....D.D.D.....	TSP-1 and 2 Consensus

Alignment of the carboxyl-terminal of the Xenopus thrombospondin-4 sequence with the last 227 amino acids of human thrombospondin-1 reveals that 60.8% of the amino acids are identical and no insertions or deletions are required. SEQ ID NO.: 2 extends 15 amino acids beyond the stop codon for human thrombospondin-1.

A particularly preferred embodiment of a thrombospondin-4 molecule has the nucleotide sequence shown in SEQ ID NO.: 3. This is a human homolog of the Xenopus sequence containing about 45 more amino acids at the amino-terminal end than the Xenopus sequence of SEQ ID NO.: 2. Approximately the first 10 nucleotides in SEQ ID NO.: 3 are linkers from the cloning library and are not thrombospondin-4 sequence. An open reading frame that is about 900 amino acids long (SEQ ID NO.: 4) is predicted from the nucleotide sequence of this human homolog.

It is not yet proved that the methionine at the 5' end of SEQ ID NO.: 3 and 4 is the beginning of the coding region. The methionine is close to the 5' end and the sequence that follows represents a reasonable signal sequence. Nevertheless, the molecular architecture of the human homolog is substantially identical to that of Xenopus

thrombospondin-4. That is, the human homolog nucleotide sequence has the same structure as Xenopus thrombospondin-4 (i.e. lacks type 1 repeats and procollagen homology, has four, type 2 repeats, and has seven, type 3 repeats).

The pattern of expression of thrombospondin-4 in human tissues is markedly different from the pattern of expression of thrombospondin 1 and 2 in human tissues. Northern blots of poly A+ selected RNA from adult human tissues was performed and probed with Xenopus thrombospondin-4 and the human homolog of Xenopus thrombospondin-4. Thrombospondin-4 showed high levels of expression in human heart and skeletal muscle (Example 3). No expression was detected in the placenta, liver or kidney. Thrombospondin-3 had its strongest Northern blot signal in the lung. The adult lung also produced the strongest signal when a blot was probed with thrombospondin-1 (Example 3). Thus, the tissue distribution of thrombospondin-4 appears to be quite different from thrombospondins 1 and 3.

Using the nucleotide sequence information provided in SEQ ID NO.: 1 and 3, cell lines expressing the thrombospondin-4 proteins can be established (Example 4). Likewise, homologs to SEQ ID NO.: 3 of other vertebrate (i.e., mammalian) species can be identified using conventional techniques, described in greater detail below. Such genetic engineering techniques are well within the scope of those of ordinary skill in the art.

The human gene encoding thrombospondin-4 has been cloned, isolated and expressed. A general protocol is present below. This protocol is intended to obtain a cDNA having a complete reading frame for the human homolog of Xenopus laevis thrombospondin-4. This objective is achieved by generating a probe to the human homolog, screening a human cDNA library with the probe and, finally, generating a coding sequence from the sequence identified in the library.

A. Cloning Xenopus laevis Thrombospondin

A cDNA encoding Xenopus thrombospondin-4 was cloned by first isolating mouse thrombospondin-1, chicken thrombospondin-2 and Xenopus thrombospondin-1 clones by screening libraries with existing probes for other species at low stringency. The resulting sequences for these thrombospondin members were aligned with human thrombospondin-1 and highly conserved regions were identified. Based on these sequences, degenerate oligonucleotides were synthesized and used as primers for the polymerase chain reaction (PCR) (SEQ ID NO.: 5 and 6; Example 1A).

The preferred primer sequences fall in the type 3 repeat domain and the carboxyl terminus of the molecule. SEQ ID NO.: 5 depicts the sequence of the forward primers and SEQ ID NO.: 6 depicts the sequence of the reverse primers.

Polymerase chain reaction (PCR) was run using Xenopus laevis cDNA as a template. PCR products were sized, fractionated and subcloned into plasmid vectors. To complete the sequence and establish the validity of the Xenopus thrombospondin-4 clone, the Xenopus cDNA library was screened using the PCR products as the probes. The probes were labelled and hybridization performed. Plaques were purified and amplified to yield high titre plate stocks. Restriction fragments were then subcloned. Sequencing was then performed using well known methods (e.g.,. chain termination method: Sanger et al., see Example 1B).

Xenopus laevis clones (designated XS3 and XS9: see Example 1B) were used to determine the nucleotide sequence of Xenopus thrombospondin-4 on both strands. Since XS9 is still 650 bp smaller than the message size predicted by Northern Blot analysis, two approaches were used to complete the sequence: (i) the Xenopus cDNA library was rescreened; and (ii) two PCR primers that include sequences within the 5' end have were used in conjunction with two PCR primers from the polylinker to perform PCR on the library. The PCR protocol

was that described in Example 1A. Neither approach yielded any additional sequences.

B. Cloning the Gene for Human Thrombospondin-4

The approach used to screen a DNA library for the presence of a thrombospondin-4 coding sequence corresponding to a human homolog includes generating preferred probes using the polymerase chain reaction. The probes were produced by using a human heart cDNA library as a template for primers (SEQ ID NO.: 7 and 8). Based on the degree of codon degeneracy of the predicted amino acid sequence, primers were derived from the Xenopus thrombospondin-4 sequence of SEQ ID NO.: 1 and 2.

The product of the PCR reaction was cloned and the human heart cDNA library rescreened using the PCR product as the probe(s) (Example 3). This preferred method required identifying tissue that expresses thrombospondin-4 as a source of RNA (e.g.. human heart tissue).

Other tissues expressing the human homolog can, however, be identified by RNA analysis, i.e.,. Northern analysis under low stringency conditions. Confirmation of a human tissue as an RNA source and identification of additional sources of tissue can be accomplished by preparing RNA from the selected tissue and performing Northern Blot Analysis under low stringency conditions using PCR product as the probe(s). A suitable range of such stringency conditions is described in Krause, M.H.. and Aaronson, S.A., 1991, Methods in Enzymology 200: 546-556. Additionally, genomic libraries can be screened for the presence of the human homolog coding sequence using a PCR generated probe(s).

C. Testing and Cloning Related Thrombospondin-4 Molecules

The invention also pertains to a more general protocol for isolating the gene for thrombospondin-4 from vertebrates, in particular from non-human vertebrates such as cows, pigs,

monkeys and the like. In this approach, total mRNA can be isolated from mammalian tissues or from cell lines likely to express thrombospondin-4 (e.g., cow or chimpanzee, heart muscle). In general, total RNA from the selected tissue or cell culture is isolated using conventional methods. Subsequent isolation of mRNA is typically accomplished by oligo (dT) chromotography. RNA for Northern analysis is size-fractionated by electrophoresis and the RNA transcripts are transferred to nitrocellulose according to conventional protocols (Sambrook, J. et al., Molecular Cloning, Cold Spring Harbor Press, N.Y.).

A labelled PCR-generated probe capable of hybridizing with the human homolog of Xenopus thrombospondin-4 (SEQ ID NO.: 3) can serve to identify RNA transcripts complementary to at least a portion of the human thrombospondin-4 gene. For example, if Northern analysis indicates that RNA isolated from a cow heart muscle hybridizes with the labelled probe, then a cow heart muscle cDNA library is a likely candidate for screening and identification of a clone containing the coding sequence for a cow homolog of thrombospondin-4.

Northern analysis is used to confirm the presence of mRNA fragments which hybridize to a probe corresponding to all or part of thrombospondin-4. Northern analysis indicates the presence and size of the transcript. This allows one to determine whether a given cDNA clone is long enough to encompass the entire transcript or whether it is necessary to obtain further cDNA clones, i.e.,. if the length of the cDNA clone is less than the length of RNA transcripts as seen by Northern analysis. If the cDNA is not long enough, it is necessary to perform several steps such as: (i) rescreen the same library with the longest probes available to identify a longer cDNA; (ii) screen a different cDNA library with the longest probe; and (iii) prepare a primer-extended cDNA library using a specific nucleotide primer corresponding to a region close to, but not at, the most 5' available region. This nucleotide sequence is used to prime reverse

transcription. The primer extended library is then screened with the probe corresponding to available sequences located 5' to the primer. See for example, Rupp et al., Neuron, 6: 811-823 (1991).

The preferred clone of thrombospondin-4 has a complete coding sequence, i.e.,. one that begins with methionine, ends with a stop codon, and preferably has another in-frame stop codon 5' to the first methionine. It is also desirable to have a cDNA that is "full length", i.e. includes all of the 5' and 3' untranslated sequences. To assemble a long clone from short fragments, the full-length sequence is determined by aligning the fragments based upon overlapping sequences. Thereafter, the full-length clone is prepared by ligating the fragments together using appropriate restriction enzymes.

As discussed above, PCR-generated probes can be used in the protocol for isolating non-human mammalian homologs to thrombospondin-4. Moreover, probes to be used in the general method for isolating non-human, vertebrate thrombospondin-4 can now include oligonucleotides, all of which are part of the human homolog shown in SEQ ID NO.: 3. Moreover, antibodies reactive with this human homolog can also be used. Unlike the PCR approach to generating a probe, the above-identified probes do not require prior isolation of RNA from a tissue expressing the vertebrate homolog.

In particular, an oligonucleotide probe typically has a sequence somewhat longer than that used for the PCR primers. A longer sequence is preferable for the probe, and it is important that codon degeneracy be minimized. A representative protocol for the preparation of an oligonucleotide probe for screening a cDNA library is described in Sambrook, J. et al. Molecular Cloning, Cold Spring Harbor Press, New York, 1989. In general, the probe is labelled, e.g.. P-32, and used to screen clones of a cDNA or genomic library.

Alternately, the library can be screened using conventional immunization techniques, such as those described

in Harlowe and Lane, D. (1988), Antibodies, Cold Spring Harbor Press, New York. Antibodies prepared using purified thrombospondin-4 as an immunogen are preferably first tested for cross reactivity with the homolog of thrombospondin-4 from other species. Other approaches to preparing antibodies for use in screening DNA libraries, as well as for use in diagnostic and research applications, are described below.

D. Nucleic Acid and Protein Sequences

The nucleic acid sequence of the human thrombospondin-4 is depicted in SEQ ID NO: 3. This sequence, its functional equivalent, or unique fragments of this sequence may be used in accordance with the invention. The term "unique fragments" refers to portions of the thrombospondin-4 nucleic acid sequence that find no counterpart in the known sequences of thrombospondins -1 and -2. Subsequences comprising hybridizable portions of the thrombospondin-4 sequence have use, e.g., in nucleic acid hybridization assays, Southern and Northern blot analyses, etc.

Nevertheless, the nucleic acid sequence depicted in SEQ ID NO: 3 can be altered by mutations such as substitutions, additions or deletions that provide for functionally equivalent nucleic acid sequences. According to the present invention, a nucleic acid sequence is "functionally equivalent" compared with the nucleic acid sequence depicted in SEQ ID NO: 3, if it satisfies at least one of the following conditions: (i) the nucleic acid sequence has the ability to hybridize to thrombospondin-4, but it does not necessarily hybridize to thrombospondin-4 with an affinity that is the same as that of the natural thrombospondin-4 nucleic acid sequence; and/or (ii) the nucleic acid can serve as a probe to distinguish between thrombospondin-4 and the other known thrombospondins. A probe that can "distinguish" between thrombospondin-4 and the other thrombospondins refers to a probe that will hybridize to a thrombospondin nucleic acid sequence that encodes for a polypeptide having has at

least four, type 2 domains; that lacks any type 1 domains and/or that lacks a region of procollagen homology. The term "probe",therefore, refers to a ligand of known qualities that can bind selectively to a target. As applied to the nucleic acid sequences of the invention, the term "probe" refers to a strand of nucleic acid having a base sequence complementary to a target strand.

Because the nucleic acid sequence of thrombospondin-4 is now known, those of ordinary skill in the art can readily determine those nucleic acid sequences of thrombospondin-4 that are not homologous to any other nucleic acid sequence, including the other thrombospondin sequences. These non-homologous sequences, and peptides encoded by them, are referred to as "unique" fragments and are meant to be included within the scope of the present invention.

Moreover, due to the degeneracy of nucleotide coding sequences, other nucleic acid sequences may be used in the practice of the present invention. These include, but are not limited to, sequences comprising all or portions of the thrombospondin-4 genes depicted in SEQ ID NO: 1 and 3 which are altered by the substitution of different codons that encode the same amino acid residue within the sequence, thus producing a silent change. Such altered sequences are regarded as equivalents of the specifically claimed sequences.

Thrombospondin-4 proteins or unique fragments or derivatives thereof include, but are not limited to, those containing as a primary amino acid sequence all, or unique parts of the amino acid residues substantially as depicted in SEQ ID NO.: 2 and SEQ ID NO.: 4, including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence, resulting in a silent change. According to the invention, an amino acid is "functionally equivalent" compared with the sequences depicted in SEQ ID NOS.: 2 and 4 if the amino acid sequence contains one or more amino acid residues within the sequence

which can be substituted by another amino acid of a similar polarity which acts as a functional equivalent. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. The non-polar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include asparatic acid and glutamic acid.

Also included within the scope of the invention are thrombospondin-4 proteins or unique fragments or derivatives thereof which are differentially modified during or after translation, e.g., by phosphorylation, glycosylation, crosslinking, acylation, proteolytic cleavage, linkage to an antibody molecule, membrane molecule or other ligand, (Ferguson et al., 1988, Ann. Rev. Biochem. 57:285-320).

In addition, the recombinant thrombospondin-4- encoding nucleic acid sequences of the invention may be engineered so as to modify processing or expression of thrombospondin-4. For example, and not by way of limitation, the thrombospondin-4 gene may be combined with a promoter sequence and/or a ribosome binding site using well characterized methods, and thereby facilitate harvesting or bioavailability.

Additionally, a given thrombospondin-4 can be mutated in vitro or in vivo, to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further in vitro modification. Any technique for mutagenesis known in the art can be used including, but not limited to, in vitro site-directed mutagenesis (Hutchinson, et al., 1978, J. Biol. Chem. 253:6551), use of TAB® linkers (Pharmacia), PCR-directed mutagenesis, and the like.

The thrombospondin-4 of the invention also includes non-human homologs of the amino acid sequence of SEQ ID NO: 4. The thrombospondin-4 peptides of the invention may be prepared by recombinant nucleic acid expression techniques or by chemical synthesis using standard peptide synthesis techniques.

Also within the scope of the invention are nucleic acid sequences or proteins encoded by nucleic acid sequences derived from the same gene but lacking one or more structural features (for instance the type 2 or 3 domains) as a result of alternative splicing of transcripts from a gene that also encodes the complete thrombospondin-4 gene, as defined previously.

Nucleic acid sequences complementary to DNA or RNA sequences encoding thrombospondin-4 or a functionally active portion thereof are also provided. In animals, particularly transgenic animals, RNA transcripts of a desired gene or genes may be translated into polypeptide products having a host of phenotypic actions. In a particular aspect of the invention, antisense thrombospondin-4 oligonucleotides can be synthesized. These oligonucleotides may have activity in their own right, such as antisense reagents which block translation or inhibit RNA function. Thus, where thrombospondin-4 is to be produced utilizing the nucleotide sequences of this invention, the DNA sequence can be in an inverted orientation which gives rise to a negative sense ("antisense") RNA on transcription. This antisense RNA is not capable of being translated to the desired thrombospondin-4 product, as it is in the wrong orientation and would give a nonsensical product if translated.

E. Expression of Thrombospondin-4

The present invention also permits the expression, isolation, and purification of the thrombospondin-4 polypeptide. A thrombospondin-4 gene may be cloned or subcloned using any method known in the art. A large number

of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, cosmids, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Viral vectors include, but are not limited to, vaccinia virus, or lambda derivatives. Plasmids include, but are not limited to, pBR322, pUC, or Bluescript® (Stratagene) plasmid derivatives. Recombinant thrombospondin-4 molecules can be introduced into host cells via transformation, transfection, infection, electroporation, etc.. Generally introduction of thrombospondin-4 molecules into a host is accomplished using a vector containing thrombospondin DNA under control by regulatory regions of the DNA that function in the host cell.

In a preferred method of expressing thrombospondin-4, the cDNA that corresponds to the entire coding region of human thrombospondin-4, constructed from two overlapping clones, was moved to the mammalian expression vector, pLEN-PT (See Example 4). The details of the experimental approach for transfection, selection and characterization of the expressed thrombospondin-4 protein were similar to those that have been used previously for human thrombospondin-1 (see Biochemistry, 31: 1173-1180 (1992)), the entire contents of which are incorporated herein by reference.

Once the thrombospondin-4 protein is expressed, it may be isolated and purified by standard methods including chromatography (e.g., ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. In particular, thrombospondin-4 protein may be isolated by binding to an affinity column comprising antibodies to thrombospondin-4 bound to a stationary support.

F. Preparation of Antibodies to Thrombospondin-4

The term "antibodies" is meant to include monoclonal antibodies, polyclonal antibodies and antibodies prepared by

recombinant nucleic acid techniques that are selectively reactive with thrombospondin-4. The term "selectively reactive" refers to those antibodies that react with thrombospondin-4, and do not react with the other thrombospondins. Antibodies include antibodies raised against Xenopus thrombospondin-4 polypeptide (SEQ ID NO.: 2) and intended to cross-react with the human homolog. These antibodies are useful for diagnostic applications. Other antibodies include antibodies raised against Xenopus thrombospondin-4, which antibodies are generally used for research purposes. These antibodies include those raised against short, synthetic peptides of the Xenopus thrombospondin-4 sequence.

Finally, antibodies are raised against the human homolog (SEQ ID NO.: 4), isolated by standard protein purification methods. Generally, a peptide immunogen is first attached to a carrier to enhance the immunogenic response. Although the peptide immunogen can correspond to any portion of the amino acid sequence of the human thrombospondin-4 protein or to variants of the sequence, such as the amino acid sequences corresponding to the primers and probes described, certain peptides are more likely than others to provoke an immediate response. For example, a peptide including the C-terminal amino acid is more likely to generate an antibody response.

Other alternatives to preparing antibodies reactive with the human homolog include: immunizing an animal with a protein expressed by a bacterial or eucaryotic cell, which cell includes the coding sequence for: (i) all or part of the human homolog; or (ii) the coding sequence for all or part of the Xenopus thrombospondin-4 protein.

Antibodies can also be prepared by immunizing an animal with whole cells that are expressing all or a part of a cDNA encoding the thrombospondin-4 protein.

To further improve the likelihood of producing an anti-thrombospondin-4 immune response, the amino acid sequence of thrombospondin-4 may be analyzed in order to

identify portions of the molecule which may be associated with increased immunogenicity. For example, the amino acid sequence may be subjected to computer analysis to identify surface epitopes which present computer-generated plots of antigenic index, an amphiphilic helix, amphiphilic sheet, hydrophilicity, and the like. Alternatively, the deduced amino acid sequences of thrombospondin-4 from different species could be compared, and relatively non-homologous regions identified. These non-homologous regions would be more likely to be immunogenic across various species.

For preparation of monoclonal antibodies directed toward thrombospondin-4, any technique which provides for the production of antibody molecules by continuous cell lines and culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (Nature, 256: 495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today, 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies, and the like, are within the scope of the present invention.

Further, single-chain antibody (SCA) methods are also available to form anti-thrombospondin-4 antibodies (Ladner et al., U.S. Patents 4,704,694 and 4,976,778).

The monoclonal antibodies may be human monoclonal antibodies or chimeric human-mouse (or other species) monoclonal antibodies. The present invention provides for antibody molecules as well as fragments of such antibody molecules.

G. Assays/Utilities

The present invention provides for assay systems in which activity or activities resulting from exposure to a peptide or non-peptide compound may be detected by measuring a physiological response to the compound in a cell or cell line which expresses the thrombospondin-4 molecules of the invention. A "physiological response" may comprise any

biological response, including but not limited to transcriptional activation of certain nucleic acid sequences (e.g., promoter/enhancer elements as well as structural genes), translation, or phosphorylation, the induction of secondary processes, and morphological changes, such as neurite sprouting.

The present invention thus provides for the development of novel assay systems which may be utilized in the screening of compounds. Target cells expressing thrombospondin-4, which bind to the compounds, may be produced by transfection with thrombospondin-4-encoding nucleic acid.

Once target cell lines are produced or identified, it may be desirable to select for cells which are exceptionally sensitive to a particular compound. Such target cells may express large amounts of thrombospondin-4; target cells expressing a relative abundance of thrombospondin-4 could be identified by selecting target cells which bind to high levels of the compound, for example cells which, when incubated with a compound/tag and subjected to immunofluorescence assay, produce a relatively higher degree of fluorescence. Alternatively, cell lines which are exceptionally sensitive to a compound may exhibit a relatively strong biological response, such as a sharp increase in immediate early gene products such as c-fos or c-jun, in response to thrombospondin-4 binding. By developing assay systems using target cells which are extremely sensitive to a compound, the present invention provides for methods of screening for low levels of thrombospondin-4 activity.

In particular, using recombinant DNA techniques, the present invention provides for thrombospondin-4 target cells which are engineered to be highly sensitive to thrombospondin-4 binding compounds. For example, the thrombospondin-4 gene, cloned according to the methods set forth above, may be inserted into cells which naturally express thrombospondin-4 such that the recombinant

thrombospondin-4 gene is expressed at high levels. Since thrombospondins generally bind large amounts of calcium, cells expressing thrombospondin-4 may find used in calcium bioassay methods, particularly in clinical settings where elevated blood calcium may be indicative of parathyroid or bone dysfunction.

The present invention also provides for experimental model systems for studying the physiological role of the native thrombospondin-4. In these model systems, thrombospondin-4 protein, peptide fragment, or a derivation thereof, may be either supplied to the system or produced within the system. Such model systems could be used to study the effects of thrombospondin-4 excess or depletion. The experimental model systems may be used to study the effects of increased or decreased response to ligand in cell or tissue cultures, in whole animals, or in particular cells or tissues within whole animals or tissue culture systems, or over specified time intervals (including during embryogenesis).

In additional embodiments of the invention, a recombinant thrombospondin-4 gene may be used to inactivate the endogenous gene by homologous recombination, and thereby create a thrombospondin-4 deficient cell, tissue, or animal. For example, and not by way of limitation, a recombinant thrombospondin-4 gene may be engineered to contain an insertional mutation (e.g.. the neo gene) which, when inserted, inactivates transcription of thrombospondin-4. Such a construct, under the control of a suitable promoter operatively linked to the thrombospondin-4 gene, may be introduced into a cell by a technique such as transfection, transduction, injection, etc.. In particular, stem cells lacking an intact thrombospondin-4 gene may generate transgenic animals deficient in thrombospondin-4. In a specific embodiment of the invention (See Example 6), the endogenous thrombospondin-4 gene of a cell may be inactivated by homologous recombination with a mutant thrombospondin-4

gene to form a transgenic animal lacking the ability to express thrombospondin-4. In another embodiment, a construct can be provided that, upon transcription, produces an "anti-sense" nucleic acid sequence which, upon translation, will not produce the required thrombospondin-4 protein.

A "transgenic animal" is an animal having cells that contain DNA which has been artificially inserted into a cell, which DNA becomes part of the genome of the animal which develops from that cell. The preferred DNA encodes for thrombospondin-4 and may be entirely foreign to the transgenic animal or may be homologous to the natural thrombospondin-4 of the transgenic animal, but which is inserted into the animal's genome at a location which differs from that of the natural homolog.

In a further embodiment of the invention, thrombospondin-4 expression may be reduced by providing thrombospondin-4 expressing cells, preferably in a transgenic animal, with an amount of thrombospondin-4 anti-sense RNA or DNA effective to reduce expression of thrombospondin-4 protein.

A transgenic animal (preferably a non-human mammal) can also be provided with a thrombospondin-4 DNA sequence that also encodes a repressor protein (e.g., the E.coli lac repressor). The repressor protein can bind to a specific DNA sequence of thrombospondin-4, thereby reducing ("repressing") the level of transcription of thrombospondin-4.

Transgenic animals of the invention which have attenuated levels of thrombospondin-4 expression have general applicability to the field of transgenic animal generation, as they permit control of the level of expression of genes.

According to the present invention, thrombospondin-4 probes may be used to identify cells and tissues of transgenic animals which lack the ability to transcribe thrombospondin-4. Thrombospondin-4 expression may be evidenced by transcription of thrombospondin-4 mRNA or production of thrombospondin-4 protein, detected using probes

which can distinguish thrombospondin-4 from thrombospondins -1 and -2, as described above. One variety of probe which may be used to detect thrombospondin-4 expression is a nucleic acid probe, containing a sequence encoding for at least four, type 2 domains. Alternatively, the probe can contain a thrombospondin sequence of the invention lacking type 1 domains or procollagen homology. Detection of thrombospondin-4-encoding mRNA may be easily accomplished by any method known in the art, including, but not limited to, in situ hybridization, Northern blot analysis, or PCR related techniques.

Another variety of probe which may be used is anti-thrombospondin-4 antibody.

The above-mentioned probes may be used experimentally to identify cells or tissues which hitherto had not been shown to express thrombospondin-4. Furthermore, these methods may be used to identify the expression of thrombospondin-4 by aberrant tissues, such as malignancies.

The invention will be further illustrated by the following, non-limiting examples.

EXAMPLE 1: Cloning the Xenopus thrombospondin-4 gene

A: Polymerase Chain Reaction

Aliquots (1,5 and 25µl) of a Xenopus laevis stage 45 cDNA library (unpublished) were brought to a final volume of 71.5µl with H_2O. The samples were heated to 70°C for 5 minutes than cooled on ice. To each sample, 10µl of 10x reaction buffer (Cetus), 6µl of 25 mM $MgCl_2$, 16µl of dNTPs and 300 pmoles of primer were added (SEQ ID NO.: 5 and 6).

The reaction mixture was heated to 95°C for 5 minutes and then equilibrated to the annealing temperature (37-48°C). TAQ polymerase (2.5 units) was added and the sample was heated to 72°C for 3 minutes. The amplification cycles were (1) incubate at 94°C for 1 minute and 20 seconds, (2)

incubate at 48°C for 2 minutes, (3) ramp to 72°C over 2 minutes, and (4) incubate at 72°C for 3 minutes. This cycle was repeated 30-40 times; finally the sample was incubated at 72°C for 7 minutes. The PCR products were separated by agarose gel electrophoresis and the appropriately sized products were subcloned into pBluescript KS or SK (Stratagene, LaJolla, CA).

B: Cloning and Sequencing

To establish the validity of the thrombospondin-4 clone and to complete the sequence, the Xenopus laevis stage 45 library was screened with the PCR product as the probe. The probe was labeled with digoxigenin-dUTP, and hybridization performed using the Genius Kit® following the supplier's protocols (Boehringer Mannheim, Indianapolis, IN). Positive plaques were taken through successive rounds of screening with the same probe at progressively lower plaque densities. The purified plaques were amplified to yield high titre plate stocks.

Because the Xenopus laevis library can be constructed in the λZAPII vector pBluescript II SK, the inserts are excised with helper phage and grown up directly following the supplier's protocols (Stratagene). BamHI and EcoRI fragments were subcloned into pBluescript II SK and KS. All sequencing was done by the chain termination method of Sanger et al. (1977) with Sequenase reagents (U.S.. Biochemical Corp., Cleveland, OH). The ends of all clones and subclones were sequenced with the remainder of the sequence being determined using synthetic oligonucleotides as primers. The sequence of Xenopus thrombospondin-4 was obtained on both strands.

The largest clone that we obtained from screening the Xenopus library was 2.8 kb. To complete the sequence, two oligonucleotides that corresponded to the bottom strand sequence near the 5' end were synthesized. The oligonucleotides and the pBluescript SK and primers (Stratagene, LaJolla, CA) were used as PCR primers with the library as the template.

Degenerate PCR using the Xenopus laevis stage 45 library has produced four distinct sequences that are related to the thrombospondins. Two of the four sequences correspond to the two copies of the thrombospondin-1 gene that are present in the Xenopus genome (Urry et al., supra 1991). In some cases, both copies of the gene are expressed (e.g., J. Biol Chem. 263: 5333-5340, DeSimone and Hynes, 1988). To date, the thrombospondin-1 sequences represent the majority of the products that we have obtained. However, two PCR products comprise sequences that are related to, but clearly distinct from thrombospondin-1. The sequences of these two PCR products (labeled TSP-4A and TSP-4B in FIG. 3, below) are very similar to each other suggesting that they represent the two copies of a newly identified gene in the Xenopus genome.

To establish that these two new sequences are derived from the Xenopus library, and to obtain more nucleotide sequence, a probe was prepared from the PCR product and used to screen the library. A screen of 120,000 plaques produced four positive clones that range in size from 1.7 kb to 2.3 kb (FIG. 3, XF1-XF4). As shown in FIG. 3, the restriction maps of the clones indicate that two distinct gene products can be identified. The longest clone for each gene (XF1 and XF3) has been sequenced on both strands. The sequence of the PCR products is included in the sequences of these clones. These data confirm that the PCR product is derived from the Xenopus library and not from another contaminating source.

When clone XF3 was used to probe a Northern blot of Xenopus stage 17 RNA, a 3.3. kb band was observed (FIG. 4). Since the message size is greater than the length of clone XF3 and the reading frame is open at the 5' end of the predicted amino acid sequence for clone XF3, the library was rescreened with the EcoRI fragment of clone XF1 in a second round of screening. This screen produced six additional clones (XS5-XS10, FIG. 3). Clone XS9 has been sequenced on both strands. Clone XS9 is approximately 469 nucleotide smaller than the message and the reading frame is open at the

5' end of the predicted amino acid sequence. The library was rescreened in a third round of screening with the EcoRI to BamHI fragment of XS9. Four additional clones have been isolated (XT11-XT14) however, they did not contain additional nucleotide sequence. To obtain additional 5' end sequences, a Xenopus laevis stage 22 library (a gift of Dr. Douglas Melton) was screened. Restriction endonuclease mapping indicated that one of the clones (XM15; not shown in Fig. 3) contained additional 5' end sequence for the TSP-4B family. A single reading frame exists between nucleotides 103 and 2970 (SEQ ID NO.: 1). There is a short (140 bp) 3' untranslated region that ends with a continuous series of adenosines. An AATAAA consensus polyadenylation signal is observed upstream of the poly A+ sequence.

Example 2: Isolating the human homolog of Xenopus thrombospondin-4

The cloning and nucleotide sequencing of Xenopus laevis thrombospondin-4 is described above. The predicted amino acid sequence (SEQ ID NO.: 2) has been searched to identify regions where the codon degeneracy is low. Two regions have been identified and the 89PCR (AAT GAG CAG GAC AAC TGT GT: SEQ ID NO.: 7) and 90PCR (TGC TCA GTC TGC TTC CAC AT: SEQ ID NO.: 8) oligonucleotides have been constructed.

Northern blot analysis of eight adult human tissues indicated that thrombospondin-4 is expressed in high levels in the heart and skeletal muscle (Example 3A). A heart cDNA library (the generous gift of Dr. Paul Allen) has been used as the template for polymerase chain reaction (PCR) with the primers 89PCR (SEQ ID NO.: 7) and 90PCR (SEQ ID NO.: 8). The product of the PCR reaction has been cloned into pBluescript vectors (Stratagene). After nucleotide sequencing to confirm that the PCR product corresponds to a sequence similar to Xenopus thrombospondin-4, the library has been screened with the PCR product as the probe. Clones have been isolated and

characterized in terms of the sites for endonuclease and nucleotide sequence. The longest clone is approximately 2kb. Computer-assisted progressive sequence alignment has been used to construct a phylogenetic tree of the thrombospondin sequences. The results of this analysis are consistent with the hypothesis that the clones that have been isolated from the human heart library represent the human homolog of Xenopus thrombospondin-4.

Example 3: Tissue Distribution of Thrombospondin-4

A. Northern Blot Analysis (General Protocol)

The Xenopus thrombospondin-4 clone XF3 was digested with EcoRI and XhoI and the insert purified. A variety of probes were used in the Northern analysis.

A human thrombospondin-1 probe was the human full-length cDNA (Lawler et al., 1992). A human thrombospondin-3 probe was developed as follows: A genomic clone GPEM-2 containing human thrombospondin-3 was kindly provided by Dr.Sandra Gendler (Imperial Cancer Research Fund, London; Lancaster et al., Biochem. Biophys. Res. Comm., 173: 1019-1-29 1990). BamHI fragments of GPEM-2 were subcloned into pBluescript KS and the ends of each clone were sequenced. One of these clones contained sequences that were homologous to the 3' end of thrombospondin-1, 2 and 4. Based on this homology, the position of the 5' end of the last exon was determined. The 3' end of this exon was taken to be the polyadenylation signal. Oligonucleotides that primed at the 5' and 3' ends of the last exon were used to amplify and clone a 293 bp DNA segment that corresponds to the last exon of human thrombospondin-3.

A third probe was a ß-actin probe (Clontech, Palo Alto, CA). The PCR product for the last exon of thrombospondin-3 and the actin probe were radiolabeled directly with the Multiprime DNA Labelling System (Amersham, Arlington Heights, IL).

A Northern blot that was prepared with Poly A+ RNA from adult human heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas was obtained from Clontech. The blot was prehybridized and hybridized as described previously (Lawler and Hynes, supra 1986).

B. Distribution of Thrombospondin-4 in Adult Human Tissues

A Northern blot of poly A+ selected RNA from eight adult human tissues is shown in FIG. 5. The lanes are represented as: (a) adult human heart; (b) adult human brain; (c) adult human placenta; (d) adult human lung; (e) adult human liver; (f) adult human skeletal muscle; (g) adult human kidney; and (h) adult human pancreas. The size of the human thrombospondin-4 message is 3.4 kb. Thrombospondin-4 (TSP-4) showed a restricted pattern of expression as this expression is visualized using a 2.2kb fragment of Xenopus thrombospondin-4. The positions and sizes of the markers are indicated on the left.

High levels of expression were observed in the heart and skeletal muscle (FIG. 5). On longer exposures, a faint band was detectable in the tissue from the brain, lung and pancreas. No expression was been detected in the placenta, liver or kidney. Comparable levels of the 2.0 kb form of ß-actin were observed in all of the lanes except the pancreas (FIG. 5). Because a considerable fraction of the total mRNA in the pancreas encodes preproinsulin and α-amylase, other mRNAs give a lower hybridization signal. Thus, although the thrombospondin-4 signal is weak in the pancreas, the relative level of expression may be significant.

When the same blot is probed for thrombospondin-3 (TSP-3), the strongest signal was observed in the lung (FIG. 5). The size of the thrombospondin-3 message was also 3.4 kb. Lower levels of thrombospondin-3 expression were observed in most of the lanes with the brain displaying the weakest hybridization signal (FIG. 5). The adult lung tissue

also produced the strongest signal when the blot was probed with a human thrombospondin-1 probe (FIG. 5; TSP-1). Varying levels of thrombospondin-1 were observed in all of the tissues on the blot. In this case, the principal message was 6.0 kb with faint bands at 4.5 and 3.6 kb.

In addition, when a Northern blot was probed with one of the clones that has been isolated from the human heart library (D7492 #9), the tissue distribution is identical to that observed when the Northern blot is probed with the Xenopus probe. These data indicate that the clones that have been isolated correspond to human thrombospondin-4. Since the Northern blot indicated that the message for human thrombospondin-4 is 3.4 kb, we rescreened the library with an approximately 450 bp EcoRI to BamHI fragment from the 5' end of the known sequence. The new clones provided additional sequence so that the total sequence is now 3074 bp. The 5' end includes a methionine residue that is followed by a 21 amino acid sequence that could represent a signal sequence.

Example 4: Expression of Thrombospondin-4

Two human thrombospondin-4 clones were used to construct a full-length coding region cDNA. An EcoRV fragment of D9892 #9 containing DNA (corresponding to nucleotides 1639 to 3074 of SEQ ID NO. 3) was cloned into EcoRv cut D9892 #11 containing DNA (corresponding to nucleotides 1 to 1638 of SEQ ID NO. 3). DNA was made from transformants and was cut with EcoRI to determine the orientation of the inserted DNA. Since the insert co-electrophoresed with the vector, the DNA was cut with XmnI followed by EcoRI to purify a full-length cDNA for thrombospondin-4 that was cloned into the EcoRI site of pLEN-PT.

The final form of each construct is moved from M13mp8 to the mammalian expression vector pLEN-PT using XbaI sites. This vector was constructed by Drs. Paul Johnson and Richard Hynes by cloning the polylinker from the pECE vector into the

BamHI site of pLEN (California Biotechnology Inc., Mountain View, CA).

Expression of the inserted DNA is driven by the human metallothionein II promoter. A mixture of the construct (5-10 µg) and neomycin resistance-containing plasmid pSV2neo (0.5-1.0 ug) is transfected into NIH 3T3 mouse fibroblast cells using the Lipofectin (Bethesda Research Laboratories, Gaithersburg, MD) protocol.

The cells are grown in 100-mm dishes until they are approximately 50% confluent. The cells are washed once with 3 mL of OptiMEMI reduced serum medium (Gibco Laboratories, Gaithersburg, M.D.) containing no serum, and then 3 mL of the same medium is placed in the dish. The DNA-Lipofectin mixture is added to the dishes with continuous swirling. After 24 h, the medium is changed to DME containing 10% FBS. After 48 h, the cells were trypsinized and replated in DME containing 10% FBS and 1 mg/mL Geneticin (G418, Gibco Laboratories). After approximately 10 days, individual G418-resistant colonies are subcloned, or the cells allowed to grow and handled as pools of G418-resistant clones. To produce culture supernatants for analysis, the cells are grown to confluence in four T75 flasks. Fresh medium is placed on the cells, and the cells are grown for 48 h. The conditioned medium is removed, and DFP added to 1 mM and PMSF added to 5 mM. After several hours at 0°C, the culture supernatants are frozen and stored at -20°C.

EXAMPLE 5: Antibody Production

A. Preparation of of Fusion Proteins

The specific methodology for construction of the fusion proteins varies depending upon the availability of restriction endonuclease sites. In general, endonuclease sites are chosen in close proximity to the region of cDNA of interest. The insert is purified by preparative agarose gel electrophoresis. The insert is isolated from the cut out

band by the glass bead method of Vogelstein and Gillespie, Proc. Nat. Acad. Sci. USA 76:615-619 (1979) or by electroelution by standard procedures recommended by the supplier (CBS Plastics). The insert is blunted and the appropriate EcoRI linker is added so that the reading frame of the insert is the same as that of the ß-galactosidase gene. The insert is cut with EcoRI and ligated into λgtll by procedures recommended by the supplier (Promega Biotec., Protoclone λgtll System). Lysogens of the Y1089 strain are selected by their ability to grow at 30°C but not at 42°C.

To prepare fusion protein, an overnight grow at 30°C is diluted 1:10 (v/v) and grown for an additional hour at 30°C. The culture is incubated at 45°C for 15 minutes and 10 µg/ml of isopropyl ß-D-thiogalactopyranoside is added. The cultures are incubated for 1 to 2 hours at 37°C. The cells are pelleted by centrifugation and resuspended in 100 mM Tris (pH 8.0), 0.25 M NaCl and 0.2 mg/ml lysozyme (Sigma). After 30 minutes at 0°C, the sample is rapidly frozen and thawed twice and then sonicated to disrupt the cells. The sample is centrifuged and the supernatant is applied to an anti-beta-galactosidase antibody affinity column (Promega Biotec, Protosorb, lacZ Immuno Affinity Adsorbent). The bound fusion protein is eluted with 0.1 M $NaHCO_3/Na_2CO_3$ (pH 10.8) and dialyzed to neutral pH.

Alternately, a glutathione S-transferase fusion protein is used as an antigen to raise a polyclonal rabbit anti-Xenopus laevis thrombospondin-4 antibody. An approximately 1.2 Kb BamHI fragment of one of the Xenopus clones (XF3) is cloned into the bacterial expression vector pGEX-2T (Pharmacia). The fusion protein is expressed and purified according to established procedures (Current Protocols in Molecular Biology, John Wiley and Sons). The fusion protein is still bound to glutathione-agarose beads when it is used as an antigen.

The antibody to human thrombospondin-4 can be produced by preparing a peptide fragment of human thrombospondin-4

believed to be immunogenic. A preferred sequence is the sequence of the last 14 amino acids that is predicted from the cDNA sequence of SEQ ID NO.: 10 (FQEFQTQNFDRFDN). This peptide is synthesized, purified, coupled to a carrier and used to produce a polyclonal antiserum in rabbits using well known methods.

B. Production of Anti-Fusion Protein Antibodies

Polyclonal rabbit antisera is produced in New Zealand White rabbits by subcutaneous injections at multiple sites of purified fusion proteins, emulsified with an equal volume of Freund's complete adjuvant. The rabbits will receive a subcutaneous booster injection after 4-6 weeks of purified antigen emulsified in Freund's incomplete adjuvant and are boosted once each month until a good titre of antibody is obtained. Rabbits are bled 10 days after boosting.

EXAMPLE 6: Preparation of Constructs for Transfections and Microinjections

Methods for purification of DNA for microinjection are well known to those of ordinary skill in the art See, for example, Hogan et al., Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1986); and Palmer et al., Nature, 300: 611 (1982).

Construction of Transgenic Animals

A variety of methods are available for the production of transgenic animals associated with this invention. DNA can be injected into the pronucleus of a fertilized egg before fusion of the male and female pronuclei, or injected into the nucleus of an embryonic cell (e.g., the nucleus of a two-cell embryo) following the initiation of cell division (Brinster et al., Proc. Nat. Acad. Sci. USA, 82: 4438-4442 (1985)). Embryos can be infected with viruses, especially retroviruses, modified to bear thrombospondin-4 genes of the invention.

Pluripotent stem cells derived from the inner cell mass of the embryo and stabilized in culture can be manipulated in culture to incorporate thrombospondin-4 genes of the invention. A transgenic animal can be produced from such cells through implantation into a blastocyst that is implanted into a foster mother and allowed to come to term.

Animals suitable for transgenic experiments can be obtained from standard commercial sources such as Charles River (Wilmington, MA), Taconic (Germantown, NY), Harlan Sprague Dawley (Indianapolis, IN), etc. Swiss Webster female mice are preferred for embryo retrieval and transfer. $B6D2F_1$ males can be used for mating and vasectomized Swiss Webster studs can be used to stimulate pseudopregnancy. Vasectomized mice and rats can be obtained from the supplier.

Microinjection Procedures

The procedures for manipulation of the rodent embryo and for microinjection of DNA into the pronucleus of the zygote are well known to those of ordinary skill in the art (Hogan et al., supra). Microinjection procedures for fish, amphibian eggs and birds are detailed in Houdebine and Chourrout, Experientia, 47: 897-905 (1991). Other procedures for introduction of DNA into tissues of animals are described in U.S. Patent No., 4,945,050 (Sandford et al., July 30, 1990).

Transgenic Mice

Female mice six weeks of age are induced to superovulate with a 5 IU injection (0.1 cc, ip) of pregnant mare serum gonadotropin (PMSG; Sigma) followed 48 hours later by a 5 IU injection (0.1 cc, ip) of human chorionic gonadotropin (hCG; Sigma). Females are placed with males immediately after hCG injection. Twenty-one hours after hCG, the mated females are sacrificed by CO_2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts and placed in Dulbecco's phosphate buffered saline (DPSS) with 0.5% bovine

serum albumin (BSA; Sigma). Surrounding cumulus cells are removed with hyaluronidase (1 mg/ml). Pronuclear embryos are then washed and placed in Earle's balanced salt solution containing 0.5% BSA (EBSS) in a 37.5°C incubator with a humidified atmosphere at 5% CO_2, 95% air until the time of injection.

Randomly cycling adult female mice are paired with vasectomized males. Swiss Webster or other comparable strains can be used for this purpose. Recipient females are mated at the same time as donor females. At the time of embryo transfer, the recipient females are anesthetized with an intraperitoneal injection of 0.015 ml of 2.5% avertin per gram of body weight. The oviducts are exposed by a single midline dorsal incision. An incision is then made through the body wall directly over the oviduct. The ovarian bursa is then torn with watchmakers forceps. Embryos to be transferred are placed in DPBS and in the tip of a transfer pipet (about 10-12 embryos). The pipet tip is inserted into the infundibulum and the embryos transferred. After the transfer, the incision is closed by two sutures.

Transgenic Rats

The procedure for generating transgenic rats is similar to that of mice See Hammer et al., Cell, 63:1099-1112 (1990). Thirty day-old female rats are given a subcutaneous injection of 20 IU of PMSG (0.1 cc) and 48 hours later each female placed with a proven male. At the same time, 40-80 day old females are placed in cages with vasectomized males. These will provide the foster mothers for embryo transfer. The next morning females are checked for vaginal plugs. Females who have mated with vasectomized males are held aside until the time of transfer. Donor females that have mated are sacrificed (CO_2 asphyxiation) and their oviducts removed, placed in DPSS with 0.5% BSA and the embryos collected. Cumulus cells surrounding the embryos are removed with hyaluronidase (1 mg/ml). The embryos are then washed

and placed in EBSS (Earle's balanced salt solution) containing 0.5% BSA in a 37.5°C incubator until the time of microinjection.

Once the embryos are injected, the live embryos are moved to DPBS for transfer into foster mothers. The foster mothers are anesthetized with ketamine (40 mg/kg, ip) and xylazine (5 mg/kg, ip). A dorsal midline incision is made through the skin and the ovary and oviduct are exposed by an incision through the muscle layer directly over the ovary. The ovarian bursa is torn, the embryos are picked up into the transfer pipet, and the tip of the transfer pipet is inserted into the infundibulum. Approximately 10-12 embryos are transferred into each rat oviduct through the infundibulum. The incision is then closed with sutures, and the foster mothers are housed singly.

Embryonic Stem (ES) Cell Methods

Introduction of DNA into ES cells:

Methods for the culturing of ES cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as electroporation, calcium phosphate/DNA precipitation; and direct injection are well known to those of ordinary skill in the art. See, for example, Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E.J. Robertson, ed., IRL Press (1987). Selection of the desired clone of thrombospondin-4-containing ES cells is accomplished through one of several means. Although embryonic stem cells are currently available for mice only, it is expected that similar methods and procedures as described and cited here will be effective for embryonic stem cells from different species as they become available.

In cases involving random gene integration, a clone containing the thrombospondin-4 gene of the invention is co-transfected with a gene encoding neomycin resistance. Alternatively, the gene encoding neomycin resistance is physically linked to the thrombospondin-4 gene. Transfection

is carried out by any one of several methods well known to those of ordinary skill in the art (E.J. Robertson, supra). Calcium phosphate/DNA precipitation, direct injection, and electroporation are the preferred methods. Following DNA introduction, cells are fed with selection medium containing 10% fetal bovine serum in DMEM supplemented with G418 (between 200 and 500μ g/ml biological weight). Colonies of cells resistant to G418 are isolated using cloning rings and expanded. DNA is extracted from drug resistant clones and Southern blotting experiments using a transgene-specific DNA probe are used to identify those clones carrying the thrombospondin-4 sequences. In some experiments, PCR methods are used to identify the clones of interest.

DNA molecules introduced into ES cells can also be integrated into the chromosome through the process of homologous recombination. Copecchi, Science, 244: 1288-1292 (1989). Direct injection results in a high efficiency of integration. Desired clones are identified through PCR of DNA prepared from pools of injected ES cells. Positive cells within the pools are identified by PCR subsequent to cell cloning. DNA introduction by electroporation is less efficient and requires a selection step. Methods for positive selection of the recombination event (i.e., neo resistance) and dual positive-negative selection (i.e., neo resistance and gancyclovir resistance) and the subsequent identification of the desired clones by PCR have been described by Copecchi, supra and Joyner et al., Nature, 338: 153-156 (1989), the disclosures of which are incorporated herein.

Embryo Recovery and ES Cell Injection:

Naturally cycling or superovulated female mice mated with males are used to harvest embryos for the implantation of ES cells. It is desirable to use the C57BL165 strain for this purpose when using mice. Embryos of the appropriate age are recovered approximately 3.5 days after successful mating.

Mated females are sacrificed by CO_2 asphyxiation or cervical dislocation and embryos are flushed from excised uterine horns and placed in Dulbecco's modified essential medium plus 10% calf serum for injection with ES cells. Approximately 10-20 ES cells are injected into blastocysts using a glass microneedle with an internal diameter of approximately 20 μm.

Transfer of Embryos to Receptive Females:

Randomly cycling adult female mice are paired with vasectomized males. Mouse strains such as Swiss Webster, ICR or others can be used for this purpose. Recipient females are mated such that they will be at 2.5 to 3.5 days post-mating when required for implantation with blastocysts containing ES cells. At the time of embryo transfer, the recipient females are anesthetized with an intraperitoneal injection of 0.015 ml of 2.5% avertin per gram of body weight. The ovaries are exposed by making an incision in the body wall directly over the oviduct and the ovary and uterus are externalized. A hole is made in the uterine horn with a 25 gauge needle through which the blastocysts are transferred. After the transfer, the ovary and uterus are pushed back into the body and the incision is closed by two sutures. This procedure is repeated on the opposite side if additional transfers are to be made.

Identification of Transgenic Mice and Rats

Tail samples (1-2 cm) are removed from three week old animals. DNA is prepared and analyzed by Southern blot or PCR to detect transgenic founder (F_0) animals and their progeny (F_1 and F_2). In this way, animals that have become transgenic for the desired thrombospondin-4 genes are identified. Because not every transgenic animal expresses the thrombospondin-4 gene, and not all of those that do will have the expression pattern anticipated by the experimenter, it is necessary to characterize each line of transgenic

animals with regard to expression of the thrombospondin-4 in different tissues.

Production of Non-Rodent Transgenic Animals

Procedures for the production of non-rodent mammals and other animals have been discussed by others. See Houdebine and Chourrout, supra; Pursel et al., Science 244: 1281-1288 (1989); and Simms et al., Bio/Technology, 6: 179-183 (1988).

Identification of Other Transgenic Organisms

An organism is identified as a potential transgenic by taking a sample of the organism for DNA extraction and hybridization analysis with a probe complementary to the thrombospondin-4 gene of interest. Alternatively, DNA extracted from the organism can be subjected to PCR analysis using PCR primers complementary to the thrombospondin-4 gene of interest.

Example 6: Protocol for Inactivating the Thrombospondin-4 Gene

Mouse genomic clones are isolated by screening a genomic library from the D3 strain of mouse with a Xenopus thrombospondin-4 probe. Duplicate lifts are hybridized with a radiolabeled probe by established protocols (Sambrook, J. et al., The Cloning Manual, Cold Spring Harbor Press, N.Y.). Plaques that correspond to positive signal on both lifts are isolated and purified by successive screening rounds at decreasing plaque density. The validity of the isolated clones is confirmed by nucleotide sequencing.

The genomic clones are used to prepare a gene targeting vector for the deletion of thrombospondin-4 in embryonic stem cells by homologous recombination. A neomycin resistance gene (neo) with its transcriptional and translational signals, is cloned into convenient sites that are near the 5' end of the gene. This will disrupt the coding sequence of

thrombospondin-4 and allow for selection by the drug Geneticin (G418) by embryonic stem (ES) cells transfected with the vector. The Herpes simplex virus thymidine kinase (HSV-tk) gene is placed at the other end of the genomic DNA as a second selectable marker. Only stem cells with the neo gene will grow in the presence of this drug.

Random integration of this construct into the ES genome will occur via sequences at the ends of the construct. In these cell lines, the HSV-tk gene will be functional and the drug gancyclovir will therefore be cytotoxic to cells having an integrated sequence of the mutated thrombospondin-4 coding sequence.

Homologous recombination will also take place between homologous DNA sequences of the ES thrombospondin-4 genome and the targeting vector. This usually results in the excision of the HSV-tk gene because it is not homologous with the thrombospondin-4 gene.

Thus, by growing the transfected ES cells in G418 and gancyclovir, the cell lines in which homologous recombination has occurred will be highly enriched. These cells will contain a disrupted coding sequence of thrombospondin-4. Individual clones are isolated and grown up to produce enough cells for frozen stocks and for preparation of DNA. Clones in which the thrombospondin-4 gene has been successfully targeted are identified by Southern blot analysis. The final phase of the procedure is to inject targeted ES cells into blastocysts and to transfer the blastocysts into pseudopregnant females. The resulting chimeric animals are bred and the offspring are analyzed by Southern blotting to identify individuals that carry the mutated form of the gene in the germ line. These animals will be mated to determine the effect of thrombospondin-4 deficiency on murine development and physiology.

It should be understood that the preceding is merely a detailed description of certain preferred embodiments. It therefore should be apparent to those skilled in the art that

various modifications and equivalents can be made without departing from the spirit or scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: BRIGHAM AND WOMEN'S HOSPITAL, INC.
(B) STREET: 75 Francis Street
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: United States of America
(F) ZIP: 02115
(G) TELEPHONE: 617-732-5504
(H) TELEFAX: 617-732-5343

(ii) TITLE OF INVENTION: HUMAN THROMBOSPONDIN-4

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.
(B) STREET: 600 Atlantic Avenue
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: United States of America
(F) ZIP: 02210

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3 1/2 inch
(B) COMPUTER: IBM-compatible
(C) OPERATING SYSTEM: MS-DOS Version 3.3
(D) SOFTWARE: WordPerfect 5.1

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: not available
(B) FILING DATE: filed herewith

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 07/985,296
(B) FILING DATE: 04-DEC-1992

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: GATES, Edward R.
(B) REGISTRATION NUMBER: 31,615
(C) REFERENCE/DOCKET NUMBER: B0801/7005WO

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2820 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iv) ANTI-SENSE: no

(vii) ORIGINAL SOURCE
(A) ORGANISM: Xenopus laevis
(D) DEVELOPMENTAL STAGE: Stage 45 (germ line)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1

CAGCCCAAGT CCACAGTTAC GCTCTTTGGA CTTTATTCCA CCAGTGACAA CAGCAGGTTC 60
TTTGAATTCA CAGTTATGGG TCGTTTAAAC AAAGCCTCTT TACGATACCT CCGGAGTGAT 120
GGGAAGTTAC ACTCAGTCTT CTTTAATAAG CTTGACATAG CTGATGGGAA GCAGCACGCG 180
CTTCTGTTGC ACCTGAGCGG CTTACACCGG GGCGCAACGT TTGCAAAGCT CTACATAGAC 240
TGTAATCCGA CAGGTGTTGT TGAAGATCTA CCCCGGCCGT TATCAGGGAT AAGGCTCAAC 300
ACAGGGTCTG TGCACTTAAG AACACTACAG AAAAAGGGAC AGGATTCCAT GGATGAATTA 360
AAACTGGTAA TGGGAGGCAC TCTGTCCGAG GTAGGAGCAA TACAAGAATG TTTTATGCAG 420
AAAAGTGAAG CCGGACAGCA GACAGGTGAC GTCAGCAGAC AGTTGATTGG CCAGATAACC 480
CAAATGAATC AGATGCTGGG AGAGCTCCGA GATGTCATGA GACAGCAGGT GAAAGAGACC 540
ATGTTCTTGA GAAACACCAT TGCAGAATGC CAGGCCTGTG GCTTAGGTCC TGACTTCCCA 600
TTGCCAACCA AAGTTCCCCA GCGCCTAGCC ACCACTACAC CTCCAAAGCC TCGATGTGAT 660
GCAACTTCAT GTTTCAGAGG AGTGCGGTGC ATTGATACAG AGGGCGGCTT CCAATGTGGG 720
CCGTGTCCTG AAGGCTATAC AGGCAACGGT GTCATTTGTA CTGATGTGGA TGAGTGTCGG 780
TTGAATCCAT GTTTCCTTGG TGTACGTTGC ATAAACACTT CTCCGGGTTT CAAATGTGAG 840
AGCTGCCCTC CCGGGTACAC TGGATCCACA ATTCAAGGGA TTGGCATTAA CTTTGCCAAG 900
CAAAATAAGC AGGTTTGCAC AGATACCAAT GAATGTGAAA ATGGAAGAAA TGGAGGGTGT 960
ACATCCAATT CTCTTTGCAT CAATACGATG GGATCTTTCC GCTGTGGGGG CTGCAAACCT 1020
GGTTATGTCG GGGATCAAAT AAAAGGCTGC AAACCTGAAA AAAGCTGCCG TCATGGACAG 1080
AATCCGTGTC ATGCAAGTGC TCAGTGTTCA GAGGAAAAGG ACGGTGACGT AACCTGCACT 1140
TGTTCAGTCG GTTGGGCCGG CAATGGCTAC CTCTGTGGCA AAGATACTGA TATTGATGGC 1200
TACCCGGATG AAGCCCTGCC ATGTCCAGAT AAGAACTGCA AAAAGGACAA CTGTGTATAT 1260
GTTCCTAACT CGGGTCAAGA AGACACTGAT AAAGATAACA TTGGAGATGC TTGTGATGAA 1320
GATGCGGATG GAGATGGTAT CCTAAATGAG CAGGACAACT GTGTGCTGGC TGCCAACATC 1380
GATCAGAAAA ACAGTGACCA AGATATATTT GGGGACGCCT GTGACAACTG CCGCTTAACC 1440
CTCAACAATG ACCAAAGGGA CACAGACAAT GACGGGAAAG GAGATGCTTG TGACGATGAC 1500
ATGGATGGAG ATGGCATCAA GAATATCTTG GATAACTGCC AGAGAGTTCC CAATGTGGAC 1560
CAGAAAGACA AAGATGGAGA TGGAGTTGGT GATATATGTG ACAGCTGTCC TGACATCATA 1620
AATCCAAACC AGTCAGACAT TGACAATGAC CTTGTTGGAG ATTCCTGTGA TACTAACCAA 1680
GACAGCGATG GTGATGGTCA CCAGGACAGC ACAGACAACT GCCCCACAGT GATAAACAGC 1740
AACCAGCTCG ACACAGACAA GGACGGCATC GGAGATGAAT GTGACGATGA TGATGATAAC 1800
GATGGAATCC CGGATACTGT TCCTCCCGGA CCTGATAACT GTAAACTGGT TCCCAACCCA 1860
GGGCAGGAGG ATGACAACAA TGATGGAGTC GGAGACGTCT GTGAGGCCGA TTTTGACCAG 1920
GACACGGTCA TTGACCGAAT TGACGTTTGC CCTGAAAATG CAGAGATCAC CCTGACAGAT 1980
TTCAGAGCTT ATCAAACTGT AGTTCTGGAT CCCGAAGGAG ATGCCCAAAT TGATCCAAAC 2040
TGGATTGTTT TGAACCAGGG AATGGAGATT GTGCAGACGA TGAACAGTGA CCCTGGACTG 2100
GCAGTTGGTT ACACAGCATT TAATGGAGTT GATTTCGAGG GCACATTCCA CGTGAACACC 2160
ATGACGGATG ATGATTACGC TGGTTTCATC TTTGGTTATC AGGACAGTTC AAGCTTTTAT 2220
GTGGTGATGT GGAAGCAGAC TGAGCAGACT TACTGGCAGG CAACCCCCTT CAGAGCAGTT 2280
GCAGAGCCTG GAATCCAACT GAAGGCTGTG AAATCCAAGT CAGGACCCGG GGAACATCTG 2340
AGGAACGCTC TGTGGCACAC AGGAGACACC AATGATCAAG TGAGGCTGCT CTGGAAAGAC 2400
CCCAGGAATG TCGGCTGGAA AGACAAAGTC TCCTACCGCT GGTTCTTACA GCACAGGCCA 2460
CAAGTCGGCT ACATCAGAGC CAGATTTTAT GAAGGCACCG AGCTGGTGGC TGACTCTGGA 2520
GTCACTGTGG ACACCACCAT GCGAGGAGGA AGACTGGGAG TATTCTGCTT TTCACAGGAA 2580

AACATAATTT GGTCCAATCT GAAATACCGG TGTAATGATA CAATCCCAGA GGATTTCCAG 2640
GCATTTCAAG CACAACAGTT TTCCAGTTAA ACAGAACCCA CACAATATCC GGTGATTTTT 2700
TTTTGTGATT TTTTTTTTGT AGTAATATGA GAAAACGTTA TTTTCATGCA GCCTTGTTTT 2760
CTACCAACTG TACAATAATG TCTGTAAAAT AAAATGGATA CAAAAATGAG AAAAAAAAAA 2820

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 889 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: yes

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2

Gln Pro Lys Ser Thr Val Thr Leu Phe Gly Leu Tyr Ser Thr Ser Asp
1 5 10 15
Asn Ser Arg Phe Phe Glu Phe Thr Val Met Gly Arg Leu Asn Lys Ala
20 25 30
Ser Leu Arg Tyr Leu Arg Ser Asp Gly Lys Leu His Ser Val Phe Phe
35 40 45
Asn Lys Leu Asp Ile Ala Asp Gly Lys Gln His Ala Leu Leu Leu His
50 55 60
Leu Ser Gly Leu His Arg Gly Ala Thr Phe Ala Lys Leu Tyr Ile Asp
65 70 75 80
Cys Asn Pro Thr Gly Val Val Glu Asp Leu Pro Arg Pro Leu Ser Gly
85 90 95
Ile Arg Leu Asn Thr Gly Ser Val His Leu Arg Thr Leu Gln Lys Lys
100 105 110
Gly Gln Asp Ser Met Asp Glu Leu Lys Leu Val Met Gly Gly Thr Leu
115 120 125
Ser Glu Val Gly Ala Ile Gln Glu Cys Phe Met Gln Lys Ser Glu Ala
130 135 140
Gly Gln Gln Thr Gly Asp Val Ser Arg Gln Leu Ile Gly Gln Ile Thr
145 150 155 160
Gln Met Asn Gln Met Leu Gly Glu Leu Arg Asp Val Met Arg Gln Gln
165 170 175
Val Lys Glu Thr Met Phe Leu Arg Asn Thr Ile Ala Glu Cys Gln Ala
180 185 190
Cys Gly Leu Gly Pro Asp Phe Pro Leu Pro Thr Lys Val Pro Gln Arg
195 200 205
Leu Ala Thr Thr Thr Pro Pro Lys Pro Arg Cys Asp Ala Thr Ser Cys
210 215 220
Phe Arg Gly Val Arg Cys Ile Asp Thr Glu Gly Gly Phe Gln Cys Gly
225 230 235 240
Pro Cys Pro Glu Gly Tyr Thr Gly Asn Gly Val Ile Cys Thr Asp Val
245 250 255
Asp Glu Cys Arg Leu Asn Pro Cys Phe Leu Gly Val Arg Cys Ile Asn
260 265 270

Thr Ser Pro Gly Phe Lys Cys Glu Ser Cys Pro Pro Gly Tyr Thr Gly
275 280 285
Ser Thr Ile Gln Gly Ile Gly Ile Asn Phe Ala Lys Gln Asn Lys Gln
290 295 300
Val Cys Thr Asp Thr Asn Glu Cys Glu Asn Gly Arg Asn Gly Gly Cys
305 310 315 320
Thr Ser Asn Ser Leu Cys Ile Asn Thr Met Gly Ser Phe Arg Cys Gly
325 330 335
Gly Cys Lys Pro Gly Tyr Val Gly Asp Gln Ile Lys Gly Cys Lys Pro
340 345 350
Glu Lys Ser Cys Arg His Gly Gln Asn Pro Cys His Ala Ser Ala Gln
355 360 365
Cys Ser Glu Glu Lys Asp Gly Asp Val Thr Cys Thr Cys Ser Val Gly
370 375 380
Trp Ala Gly Asn Gly Tyr Leu Cys Gly Lys Asp Thr Asp Ile Asp Gly
385 390 395 400
Tyr Pro Asp Glu Ala Leu Pro Cys Pro Asp Lys Asn Cys Lys Lys Asp
405 410 415
Asn Cys Val Tyr Val Pro Asn Ser Gly Gln Glu Asp Thr Asp Lys Asp
420 425 430
Asn Ile Gly Asp Ala Cys Asp Glu Asp Ala Asp Gly Asp Gly Ile Leu
435 440 445
Asn Glu Gln Asp Asn Cys Val Leu Ala Ala Asn Ile Asp Gln Lys Asn
450 455 460
Ser Asp Gln Asp Ile Phe Gly Asp Ala Cys Asp Asn Cys Arg Leu Thr
465 470 475 480
Leu Asn Asn Asp Gln Arg Asp Thr Asp Asn Asp Gly Lys Gly Asp Ala
485 490 495
Cys Asp Asp Asp Met Asp Gly Asp Gly Ile Lys Asn Ile Leu Asp Asn
500 505 510
Cys Gln Arg Val Pro Asn Val Asp Gln Lys Asp Lys Asp Gly Asp Gly
515 520 525
Val Gly Asp Ile Cys Asp Ser Cys Pro Asp Ile Ile Asn Pro Asn Gln
530 535 540
Ser Asp Ile Asp Asn Asp Leu Val Gly Asp Ser Cys Asp Thr Asn Gln
545 550 555 560
Asp Ser Asp Gly Asp Gly His Gln Asp Ser Thr Asp Asn Cys Pro Thr
565 570 575
Val Ile Asn Ser Asn Gln Leu Asp Thr Asp Lys Asp Gly Ile Gly Asp
580 585 590
Glu Cys Asp Asp Asp Asp Asp Asn Asp Gly Ile Pro Asp Thr Val Pro
595 600 605
Pro Gly Pro Asp Asn Cys Lys Leu Val Pro Asn Pro Gly Gln Glu Asp
610 615 620
Asp Asn Asn Asp Gly Val Gly Asp Val Cys Glu Ala Asp Phe Asp Gln
625 630 635 640
Asp Thr Val Ile Asp Arg Ile Asp Val Cys Pro Glu Asn Ala Glu Ile
645 650 655
Thr Leu Thr Asp Phe Arg Ala Tyr Gln Thr Val Val Leu Asp Pro Glu
660 665 670
Gly Asp Ala Gln Ile Asp Pro Asn Trp Ile Val Leu Asn Gln Gly Met
675 680 685
Glu Ile Val Gln Thr Met Asn Ser Asp Pro Gly Leu Ala Val Gly Tyr
690 695 700

```
Thr Ala Phe Asn Gly Val Asp Phe Glu Gly Thr Phe His Val Asn Thr
705                 710                 715                 720
Met Thr Asp Asp Asp Tyr Ala Gly Phe Ile Phe Gly Tyr Gln Asp Ser
                725                 730                 735
Ser Ser Phe Tyr Val Val Met Trp Lys Gln Thr Glu Gln Thr Tyr Trp
            740                 745                 750
Gln Ala Thr Pro Phe Arg Ala Val Ala Glu Pro Gly Ile Gln Leu Lys
        755                 760                 765
Ala Val Lys Ser Lys Ser Gly Pro Gly Glu His Leu Arg Asn Ala Leu
    770                 775                 780
Trp His Thr Gly Asp Thr Asn Asp Gln Val Arg Leu Leu Trp Lys Asp
785                 790                 795                 800
Pro Arg Asn Val Gly Trp Lys Asp Lys Val Ser Tyr Arg Trp Phe Leu
                805                 810                 815
Gln His Arg Pro Gln Val Gly Tyr Ile Arg Ala Arg Phe Tyr Glu Gly
            820                 825                 830
Thr Glu Leu Val Ala Asp Ser Gly Val Thr Val Asp Thr Thr Met Arg
        835                 840                 845
Gly Gly Arg Leu Gly Val Phe Cys Phe Ser Gln Glu Asn Ile Ile Trp
    850                 855                 860
Ser Asn Leu Lys Tyr Arg Cys Asn Asp Thr Ile Pro Glu Asp Phe Gln
865                 870                 875                 880
Ala Phe Gln Ala Gln Gln Phe Ser Ser
                885
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3074 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3

```
GAATTCCGGG GAGCAGGAAG AGCCAACATG CTGGCCCCGC GCGGAGCCGC CGTCCTCCTG   60
CTGCACCTGG TCCTGCAGCG GTGGCTAGCG GCAGGCGCCC AGGCCACCCC CCAGGTCTTT  120
GACCTTCTCC CATCTTCCAG TCAGAGGCTA AACCCAGGCG CTCTGCTGCC AGTCCTGACA  180
GACCCCGCCC TGAATGATCT CTATGTGATT TCCACCTTCA AGCTGCAGAC TAAAAGTTCA  240
GCCACCATCT TCGGTCTTTA CTCTTCAACT GACAACAGTA AATATTTTGA ATTTACTGTG  300
ATGGGACGCT TAAGCAAAGC CATCCTCCGT TACCTGAAGA ACGATGGGAA GGTGCATTTG  360
GTGGTTTTCA ACAACCTGCA GCTGGCAGAC GGAAGGCGGC ACAGGATCCT CCTGAGGCTG  420
AGCAATTTGC AGCGAGGGGC CGGCTCCCTA GAGCTCTACC TGGACTGCAT CCAGGTGGAT  480
TCCGTTCACA ATCTCCCCAG GGCCTTTGCT GGCCCCTCCC AGAAACCTGA GACCATTGAA  540
TTGAGGACTT TCCAGAGGAA GCCACAGGAC TTCTTGGAAG AGCTGAAGCT GGTGGTGAGA  600
GGCTCACTGT TCCAGGTGGC CAGCCTGCAA GACTGCTTCC TGCAGCAGAG TGAGCCACTG  660
GCTGCCACAG GCACAGGGGA CTTTAACCGG CAGTTCTTGG GTCAAATGAC ACAATTAAAC  720
CAACTCCTGG GAGAGGTGAA GGACCTTCTG AGACAGCAGG TTAAGGAAAC ATCATTTTTG  780
```

```
CGAAACACCA TAGCTGAATG CCAGGCTTGC GGTCCTCTCA AGTTTCAGTC TCCGACCCCA  840
AGCACGGTGG TCGCCCCGGC TCCCCCTGCA CCGCCAACAC GCCCACCTCG TCGGTGTGAC  900
TCCAACCCAT GTTTCCGAGG TGTCCAATGT ACCGACAGTA GAGATGGCTT CCAGTGTGGG  960
CCCTGCCCCG AGGGCTACAC AGGAAACGGG ATCACCTGTA TTGATGTTGA TGAGTGCAAA 1020
TACCATCCCT GCTACCCGGG CGTGCACTGC ATAAATTTGT CTCCTGGCTT CAGATGTGAC 1080
GCCTGCCCAG TGGGCTTCAC AGGGCCCATG GTGCAGGGTG TTGGGATCAG TTTTGCCAAG 1140
TCAAACAAGC AGGTCTGCAC TGACATTGAT GAGTGTCGAA ATGGAGCGTG CGTTCCCAAC 1200
TCGATCTGCG TTAATACTTT GGGATCTTAC CGCTGTGGGC CTTGTAAGCC GGGGTATACT 1260
GGTGATCAGA TAAGGGGATG CAAAGTGGAA AGAAACTGCA GAAACCCAGA GCTGAACCCT 1320
TGCAGTGTGA ATGCCCAGTG CATTGAAGAG AGGCAGGGGG ATGTGACATG TGTGTGTGGA 1380
GTCGGTTGGG CTGGAGATGG CTATATCTGT GGAAAGGATG TGGACATCGA CAGTTACCCC 1440
GACGAAGAAC TGCCATGCTC TGCCAGGAAC TGTAAAAAGG ACAACTGCAA ATATGTGCCA 1500
AATTCTGGCC AAGAAGATGC AGACAGAGAT GGCATTGGCG ACGCTTGTGA CGAGGATGCT 1560
GACGGAGATG GGATCCTGAA TGAGCAGGAT AACTGTGTCC TGATTCATAA TGTGGACCAA 1620
AGGAACAGCG ATAAAGATAT CTTTGGGGAT GCCTGTGATA ACTGCCTGAG TGTCTTAAAT 1680
AACGACCAGA AAGACACCGA TGGGGATGGA AGAGGAGATG CCTGTGATGA TGACATGGAT 1740
GGAGATGGAA TAAAAAACAT TCTGGACAAC TGCCCAAAAT TTCCCAATCG TGACCAACGG 1800
GACAAGGATG GTGATGGTGT GGGGGATGCC TGTGACAGTT GTCCTGATGT CAGCAACCCT 1860
AACCAGTCTG ATGTGGATAA TGATCTGGTT GGGGACTCCT GTGACACCAA TCAGGACAGT 1920
GATGGAGATG GGCACCAGGA CAGCACAGAC AACTGCCCCA CCGTCATTAA CAGTGCCCAG 1980
CTGGACACCG ATAAGGATGG AATTGGTGAC GAGTGTGATG ATGATGATGA CAATGATGGT 2040
ATCCCAGACC TGGTGCCCCC TGGACCAGAC AACTGCCGGC TGGTCCCCAA CCCAGCCCAG 2100
GAGGATAGCA ACAGCGACGG AGTGGGAGAC ATCTGTGAGT CTGACTTTGA CCAGGACCAG 2160
GTCATCGATC GGATCGACGT CTGCCCAGAG AACGCAGAGG TCACCCTGAC CGACTTCAGG 2220
GCTTACCAGA CCGTGGGCCT GGATCCTGAA GGGGATGCCC AGATCGATCC CAACTGGGTG 2280
GTCCTGAACC AGGGCATGGA GATTGTACAG ACCATGAACA GTGATCCTGG CCTGGCAGTG 2340
GGGTACACAG CTTTTAATGG AGTTGACTTC GAAGGGACCT TCCATGTGAA TACCCAGACA 2400
GATGATGACT ATGCAGGCTT TATCTTTGGC TACCAAGATA GCTCCAGCTT CTACGTGGTC 2460
ATGTGGAAGC AGACGGAGCA GACATATTGG CAAGCCACCC CATTCCGAGC AGTTGCAGAA 2520
CCTGGCATTC AGCTCAAGGC TGTGAAGTCT AAGACAGGTC CAGGGGAGCA TCTCCGGAAC 2580
TCCCTGTGGC ACACGGGGGA CACCAGTGAC CAGGTCAGGC TGCTGTGGAA GGACTCCAGG 2640
AATGTGGGCT GGAAGGACAA GGTGTCCTAC CGCTGGTTCC TACAGCACAG GCCCCAGGTG 2700
GGCTACATCA GGGTACGATT TTATGAAGGC TCTGAGTTGG TGGCTGACTC TGGCGTCACC 2760
ATAGACACCA CAATGCGTGG AGGCCGACTT GGCGTTTTCT GCTTCTCTCA AGAAAACATC 2820
ATCTGGTCCA ACCTCAAGTA TCGCTGCAAT GACACCATCC CTGAGGACTT CCAAGAGTTT 2880
CAAACCCAGA ATTTCGACCG CTTCGATAAT TAAACCAAGG AAGCAATCTG TAACTGCTTT 2940
TCGGAACACT AAAACCATAT ATATTTTAAC TTCAATTTTC TTTAGCTTTT ACCAACCCAA 3000
ATATATCAAA ACGTTTTATG TGAATGTGGC AATAAAGGAG AAGAGATCAT TTTTAAAAAA 3060
AAAAAAAAAA AAAA                                                    3074
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 961 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: yes

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4

Met Leu Ala Pro Arg Gly Ala Ala Val Leu Leu Leu His Leu Val Leu
1 5 10 15
Gln Arg Trp Leu Ala Ala Gly Ala Gln Ala Thr Pro Gln Val Phe Asp
20 25 30
Leu Leu Pro Ser Ser Ser Gln Arg Leu Asn Pro Gly Ala Leu Leu Pro
35 40 45
Val Leu Thr Asp Pro Ala Leu Asn Asp Leu Tyr Val Ile Ser Thr Phe
50 55 60
Lys Leu Gln Thr Lys Ser Ser Ala Thr Ile Phe Gly Leu Tyr Ser Ser
65 70 75 80
Thr Asp Asn Ser Lys Tyr Phe Glu Phe Thr Val Met Gly Arg Leu Ser
85 90 95
Lys Ala Ile Leu Arg Tyr Leu Lys Asn Asp Gly Lys Val His Leu Val
100 105 110
Val Phe Asn Asn Leu Gln Leu Ala Asp Gly Arg Arg His Arg Ile Leu
115 120 125
Leu Arg Leu Ser Asn Leu Gln Arg Gly Ala Gly Ser Leu Glu Leu Tyr
130 135 140
Leu Asp Cys Ile Gln Val Asp Ser Val His Asn Leu Pro Arg Ala Phe
145 150 155 160
Ala Gly Pro Ser Gln Lys Pro Glu Thr Ile Glu Leu Arg Thr Phe Gln
165 170 175
Arg Lys Pro Gln Asp Phe Leu Glu Glu Leu Lys Leu Val Val Arg Gly
180 185 190
Ser Leu Phe Gln Val Ala Ser Leu Gln Asp Cys Phe Leu Gln Gln Ser
195 200 205
Glu Pro Leu Ala Ala Thr Gly Thr Gly Asp Phe Asn Arg Gln Phe Leu
210 215 220
Gly Gln Met Thr Gln Leu Asn Gln Leu Leu Gly Glu Val Lys Asp Leu
225 230 235 240
Leu Arg Gln Gln Val Lys Glu Thr Ser Phe Leu Arg Asn Thr Ile Ala
245 250 255
Glu Cys Gln Ala Cys Gly Pro Leu Lys Phe Gln Ser Pro Thr Pro Ser
260 265 270
Thr Val Val Ala Pro Ala Pro Pro Ala Pro Pro Thr Arg Pro Pro Arg
275 280 285
Arg Cys Asp Ser Asn Pro Cys Phe Arg Gly Val Gln Cys Thr Asp Ser
290 295 300
Arg Asp Gly Phe Gln Cys Gly Pro Cys Pro Glu Gly Tyr Thr Gly Asn
305 310 315 320
Gly Ile Thr Cys Ile Asp Val Asp Glu Cys Lys Tyr His Pro Cys Tyr
325 330 335
Pro Gly Val His Cys Ile Asn Leu Ser Pro Gly Phe Arg Cys Asp Ala
340 345 350
Cys Pro Val Gly Phe Thr Gly Pro Met Val Gln Gly Val Gly Ile Ser
355 360 365
Phe Ala Lys Ser Asn Lys Gln Val Cys Thr Asp Ile Asp Glu Cys Arg
370 375 380
Asn Gly Ala Cys Val Pro Asn Ser Ile Cys Val Asn Thr Leu Gly Ser
385 390 395 400
Tyr Arg Cys Gly Pro Cys Lys Pro Gly Tyr Thr Gly Asp Gln Ile Arg
405 410 415

Gly Cys Lys Val Glu Arg Asn Cys Arg Asn Pro Glu Leu Asn Pro Cys
420 425 430
Ser Val Asn Ala Gln Cys Ile Glu Glu Arg Gln Gly Asp Val Thr Cys
435 440 445
Val Cys Gly Val Gly Trp Ala Gly Asp Gly Tyr Ile Cys Gly Lys Asp
450 455 460
Val Asp Ile Asp Ser Tyr Pro Asp Glu Glu Leu Pro Cys Ser Ala Arg
365 470 475 480
Asn Cys Lys Lys Asp Asn Cys Lys Tyr Val Pro Asn Ser Gly Gln Glu
485 490 495
Asp Ala Asp Arg Asp Gly Ile Gly Asp Ala Cys Asp Glu Asp Ala Asp
500 505 510
Gly Asp Gly Ile Leu Asn Glu Gln Asp Asn Cys Val Leu Ile His Asn
515 520 525
Val Asp Gln Arg Asn Ser Asp Lys Asp Ile Phe Gly Asp Ala Cys Asp
530 535 540
Asn Cys Leu Ser Val Leu Asn Asn Asp Gln Lys Asp Thr Asp Gly Asp
545 550 555 560
Gly Arg Gly Asp Ala Cys Asp Asp Asp Met Asp Gly Asp Gly Ile Lys
565 570 575
Asn Ile Leu Asp Asn Cys Pro Lys Phe Pro Asn Arg Asp Gln Arg Asp
580 585 590
Lys Asp Gly Asp Gly Val Gly Asp Ala Cys Asp Ser Cys Pro Asp Val
595 600 605
Ser Asn Pro Asn Gln Ser Asp Val Asp Asn Asp Leu Val Gly Asp Ser
610 615 620
Cys Asp Thr Asn Gln Asp Ser Asp Gly Asp Gly His Gln Asp Ser Thr
625 630 635 640
Asp Asn Cys Pro Thr Val Ile Asn Ser Ala Gln Leu Asp Thr Asp Lys
645 650 655
Asp Gly Ile Gly Asp Glu Cys Asp Asp Asp Asp Asp Asn Asp Gly Ile
660 665 670
Pro Asp Leu Val Pro Pro Gly Pro Asp Asn Cys Arg Leu Val Pro Asn
675 680 685
Pro Ala Gln Glu Asp Ser Asn Ser Asp Gly Val Gly Asp Ile Cys Glu
690 695 700
Ser Asp Phe Asp Gln Asp Gln Val Ile Asp Arg Ile Asp Val Cys Pro
705 710 715 720
Glu Asn Ala Glu Val Thr Leu Thr Asp Phe Arg Ala Tyr Gln Thr Val
725 730 735
Gly Leu Asp Pro Glu Gly Asp Ala Gln Ile Asp Pro Asn Trp Val Val
740 745 750
Leu Asn Gln Gly Met Glu Ile Val Gln Thr Met Asn Ser Asp Pro Gly
755 760 765
Leu Ala Val Gly Tyr Thr Ala Phe Asn Gly Val Asp Phe Glu Gly Thr
770 775 780
Phe His Val Asn Thr Gln Thr Asp Asp Asp Tyr Ala Gly Phe Ile Phe
785 790 795 800
Gly Tyr Gln Asp Ser Ser Ser Phe Tyr Val Val Met Trp Lys Gln Thr
805 810 815
Glu Gln Thr Tyr Trp Gln Ala Thr Pro Phe Arg Ala Val Ala Glu Pro
820 825 830
Gly Ile Gln Leu Lys Ala Val Lys Ser Lys Thr Gly Pro Gly Glu His
835 840 845

Leu Arg Asn Ser Leu Trp His Thr Gly Asp Thr Ser Asp Gln Val Arg
850 855 860
Leu Leu Trp Lys Asp Ser Arg Asn Val Gly Trp Lys Asp Lys Val Ser
865 870 875 880
Tyr Arg Trp Phe Leu Gln His Arg Pro Gln Val Gly Tyr Ile Arg Val
885 890 895
Arg Phe Tyr Glu Gly Ser Glu Leu Val Ala Asp Ser Gly Val Thr Ile
900 905 910
Asp Thr Thr Met Arg Gly Gly Arg Leu Gly Val Phe Cys Phe Ser Gln
915 920 925
Glu Asn Ile Ile Trp Ser Asn Leu Lys Tyr Arg Cys Asn Asp Thr Ile
930 935 940
Pro Glu Asp Phe Gln Glu Phe Gln Thr Gln Asn Phe Asp Arg Phe Asp
945 950 955 960
Asn

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid
(A) DESCRIPTION: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5

GACTGAATTC CYAAYGCYAA CCAGGCHGAY CAYGAYAARG AYGG 44

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid
(A) DESCRIPTION: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6

CTAGGAATTC CTGKCCDGGR GTGTTTCCKG TRTGCCA 37

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid
(A) DESCRIPTION: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7

AATGAGCAGG ACAACTGTGT 20

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid
(A) DESCRIPTION: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8

TGCTCAGTCT GCTTCCACAT 20

CLAIMS

1. An isolated thrombospondin that has four, type 2 domains or unique fragments thereof.

2. An isolated thrombospondin that is free of type 1 domains.

3. An isolated thrombospondin that is free of regions of homology to procollagen.

4. An isolated thrombospondin that has at least four, type 2 domains, that is free of type 1 domains, and that is free of regions of homology to procollagen.

5. An isolated nucleic acid encoding a thrombospondin that has four, type 2 domains, or unique fragments thereof.

6. An isolated nucleic acid encoding a thrombospondin that is free of type 1 domains, or unique fragments thereof.

7. An isolated nucleic acid encoding a thrombospondin that is free of regions of homology to procollagen, or unique fragments thereof.

8. An isolated nucleic acid encoding a thrombospondin that has four, type 2 domains, is free of type 1 domains, and is free of regions of homology to procollagen.

9. An isolated nucleotide sequence encoding at least a portion of platelet thrombospondin, said portion having at least four, type 2 domains.

10. The isolated nucleotide sequence of claim 9, encoding an amino acid sequence selected from the group consisting of SEQ ID NOS.:2 and 4.

11. The isolated nucleotide sequence of claim 9, comprising SEQ ID NO.:1.

12. The isolated nucleotide sequence of claim 9, comprising SEQ ID NO.: 3.

13. The isolated nucleotide sequence of claim 9, said portion free of any type 1 domains.

14. The isolated nucleotide sequence of claims 9 or 13, said portion free of regions of homology to procollagen.

15. An isolated polypeptide comprising the expression product of a nucleotide sequence encoding at least a portion of a platelet thrombospondin gene, wherein said nucleotide sequence encodes four, type 2 domains.

16. The isolated polypeptide of claim 15, said nucleotide sequence lacking a sequence encoding for type 1 domains.

17. The isolated polypeptide of claim 16, said nucleotide sequence lacking a sequence encoding for regions of homology to procollagen.

18. An isolated polypeptide selected from the group consisting of SEQ ID NO.: 2 and 4.

19. A probe capable of distinguishing thrombospondin-4, from thrombospondins -1, and -2.

20. The probe of claim 19, comprising a DNA sequence having at least four, type 2 domains.

21. The probe of claims 19, comprising a DNA sequence lacking any type 1 domains.

22. The probe of claim 19, comprising a DNA sequence lacking a region of homology with procollagen.

23. A recombinant vector, said vector having a nucleotide sequence for transcription into a messenger RNA encoding a thrombospondin of claims 1, 2, 3 or 4.

24. A microorganism containing a recombinant expression vector, said vector comprising a nucleotide sequence encoding for a thrombospondin of claims 1, 2, 3 or 4.

25. A nucleic acid sequence comprising a transcriptional promoter linked to a nucleic acid sequence encoding a thrombospondin that has at least four, type 2 domains, said nucleic acid sequence in an orientation which, upon transcription, results in a negative RNA transcript.

26. The nucleic acid sequence of claim 25, said sequence free of type 1 domains.

27. The nucleic acid sequence of claim 26, said nucleotide sequence free of regions of homology with procollagen.

28. An antibody selectively reactive with thrombospondin polypeptide having at least four, type 2 domains.

29. The antibody of claim 28, said thrombospondin free of type 1 domains.

30. The antibody of claim 29, said platelet thrombospondin free of regions of homology with procollagen.

31. A method for producing platelet thrombospondin polypeptide, comprising,

introducing an expression vector into a host, said vector containing a DNA sequence encoding at least a portion of a polypeptide characterized as platelet thrombospondin, said DNA sequence containing at least four, type 2 domains, said DNA sequence under control by regulatory regions functional in said host, whereby said polypeptide is expressed;

allowing said host to express said polypeptide as an expression product; and

isolating said expression product.

32. The method of claim 31, wherein said expression vector provided to the host includes a DNA sequence selected from the group consisting of SEQ ID NO.: 1 and 3.

33. The method of claim 31, wherein the expression vector provided to the host includes a DNA sequence free of type 1 domains.

34. The method of claim 31, wherein the expression vector introduced into the host includes a DNA sequence free of regions of homology with procollagen.

35. A method for inactivating a gene for platelet thrombospondin, comprising:

providing a construct including a nucleotide sequence encoding for at least a portion of platelet thrombospondin having at least four, type 2 domains, said sequence which, when inserted, inactivates production of said platelet thrombospondin, said construct further having a promotor operatively linked to said sequence;

introducing said construct into a cell;

allowing said construct to homologously recombine with complementary sequences of said cell; and

selecting for cells lacking the ability to produce said platelet thrombospondin having at least four, type 2 domains.

36. The method of claim 35, wherein said introduced construct comprises a nucleotide sequence encoding for platelet thrombospondin that is free of type 1 domains.

37. The method of claim 36, wherein said introduced construct comprises a thrombospondin nucleotide sequence encoding for platelet thrombospondin that is free of regions of homology with procollagen.

38. The method of claim 35, wherein the step of introducing said contruct into a cell comprises introducing said construct in a mammalian stem cell.

39. A transgenic non-human vertebrate animal, all of whose cells contain a nucleotide sequence encoding for platelet thrombospondin-4.

40. The transgenic animal of claim 39, wherein said polypeptide has at least four, type 2 domains.

41. The transgenic animal of claim 39, wherein said polypeptide lacks any type 1 domains.

42. The transgenic animal of claim 39, wherein said polypeptide lacks a region of homology to procollagen.

43. A thrombospondin polypeptide expressed in heart and skeletal muscle and not expressed in placenta, liver, or kidney.

44. The polypeptide of claim 43, wherein said polypeptide has at least four, type 2 domains.

45. The polypeptide of claims 43 or 44, wherein said polypeptide lacks any type 1 domains.

46. The polypeptide of claim 45, wherein said polypeptide lacks a region of homology with procollagen.

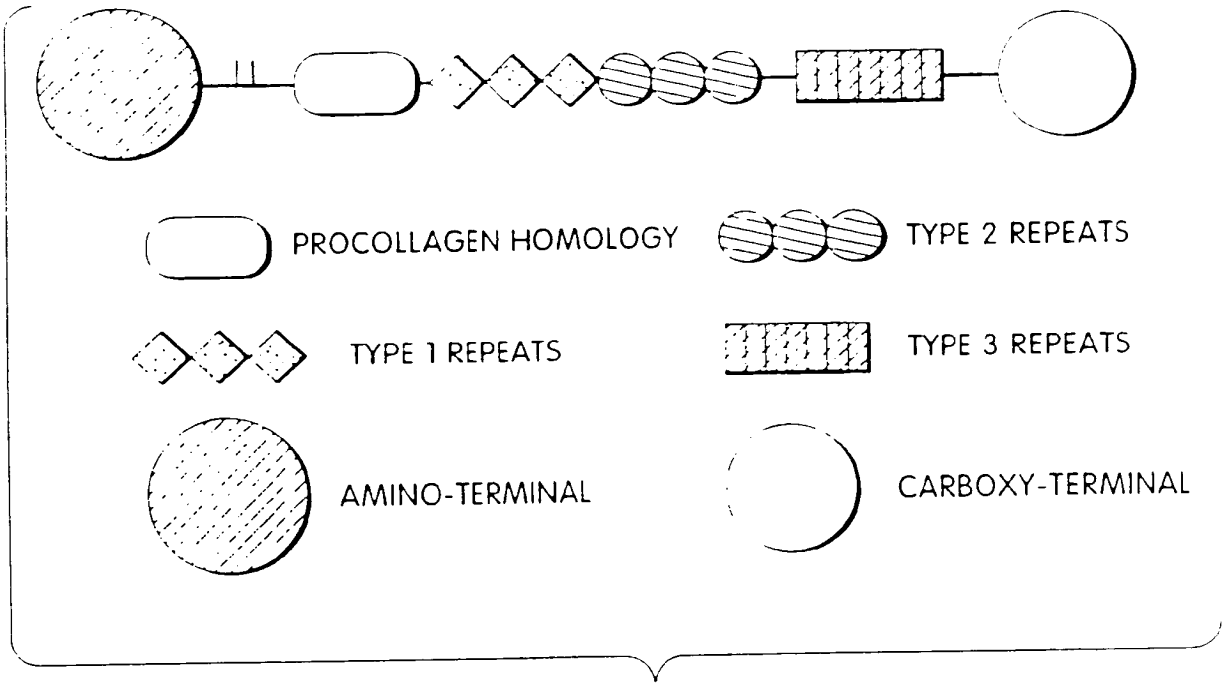

PROCOLLAGEN HOMOLOGY
TYPE 2 REPEATS
TYPE 1 REPEATS
TYPE 3 REPEATS
AMINO-TERMINAL
CARBOXY-TERMINAL

Fig. 1

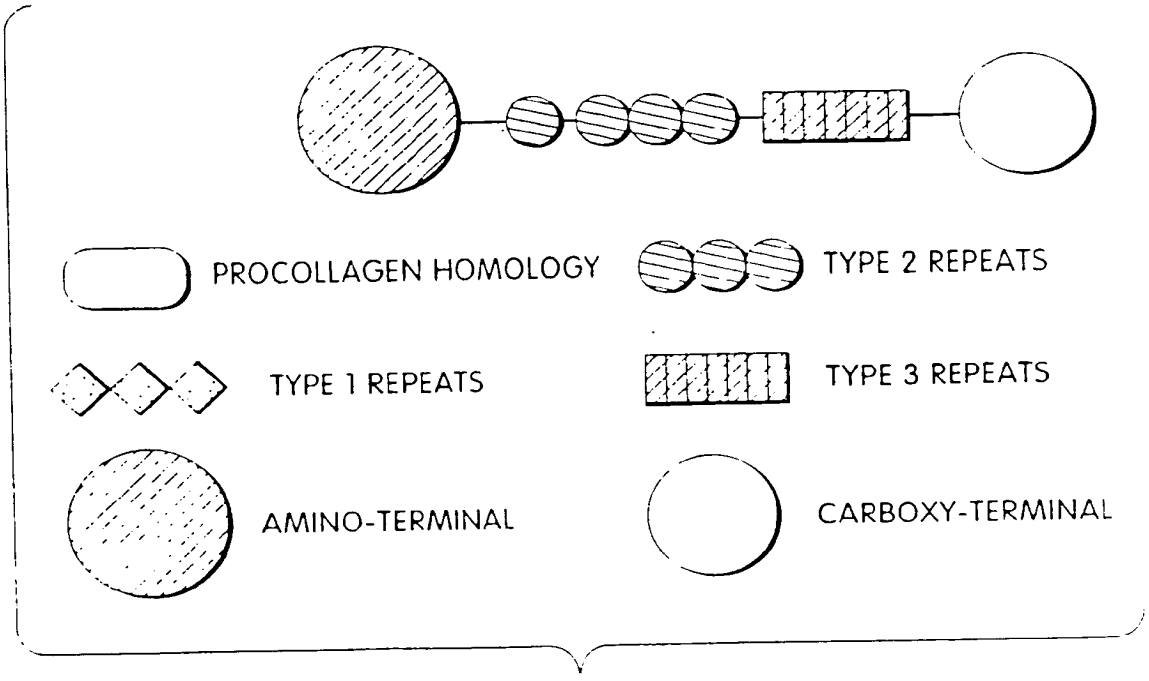

PROCOLLAGEN HOMOLOGY
TYPE 2 REPEATS
TYPE 1 REPEATS
TYPE 3 REPEATS
AMINO-TERMINAL
CARBOXY-TERMINAL

Fig. 2

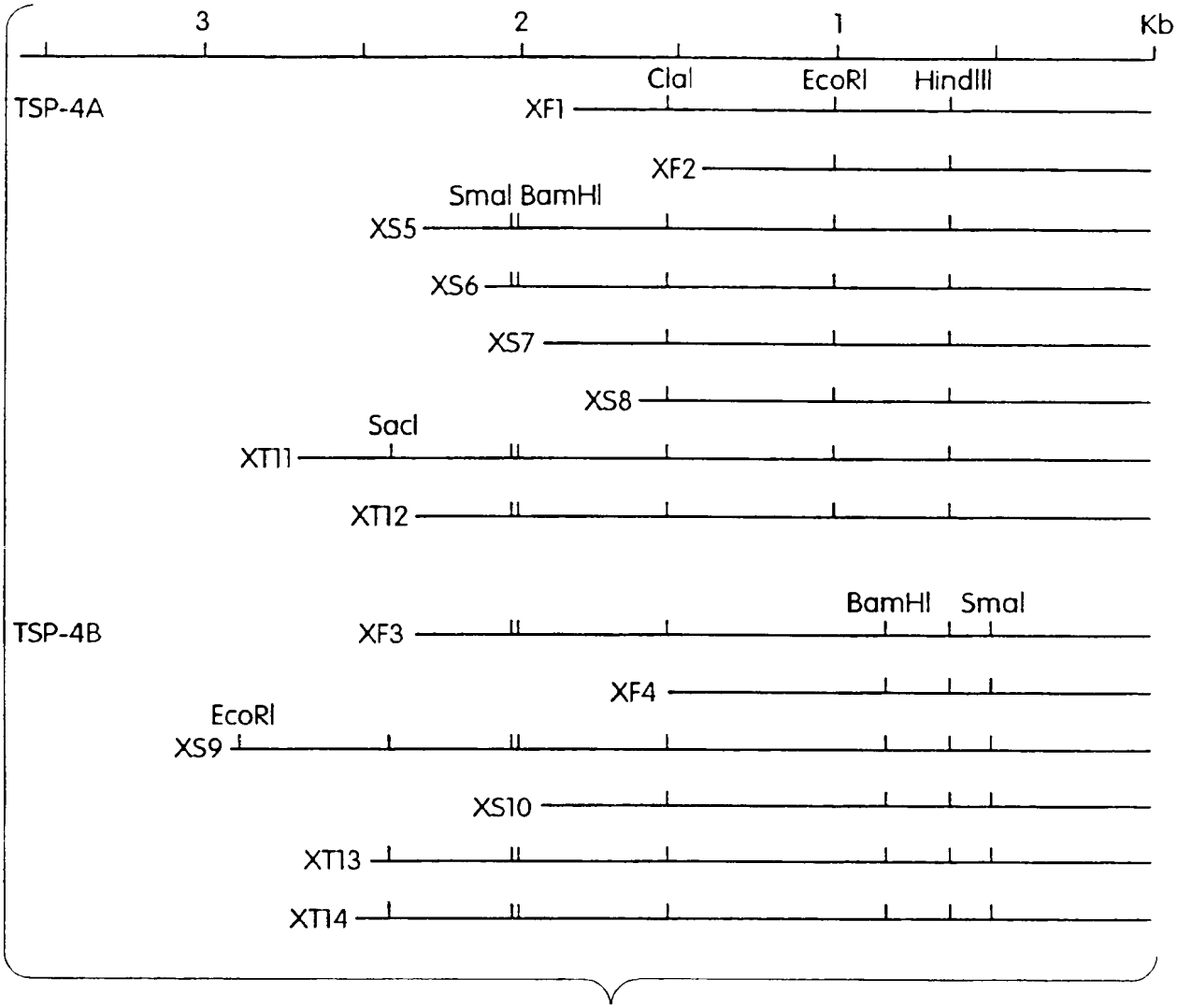

3
2
1
Kb
ClaI
EcoRI
HindIII
TSP-4A
XF1
XF2
SmaI BamHI
XS5
XS6
XS7
XS8
SacI
XT11
XT12
BamHI
SmaI
TSP-4B
XF3
XF4
EcoRI
XS9
XS10
XT13
XT14

Fig. 3

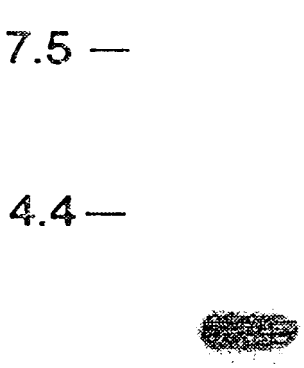


Fig. 4

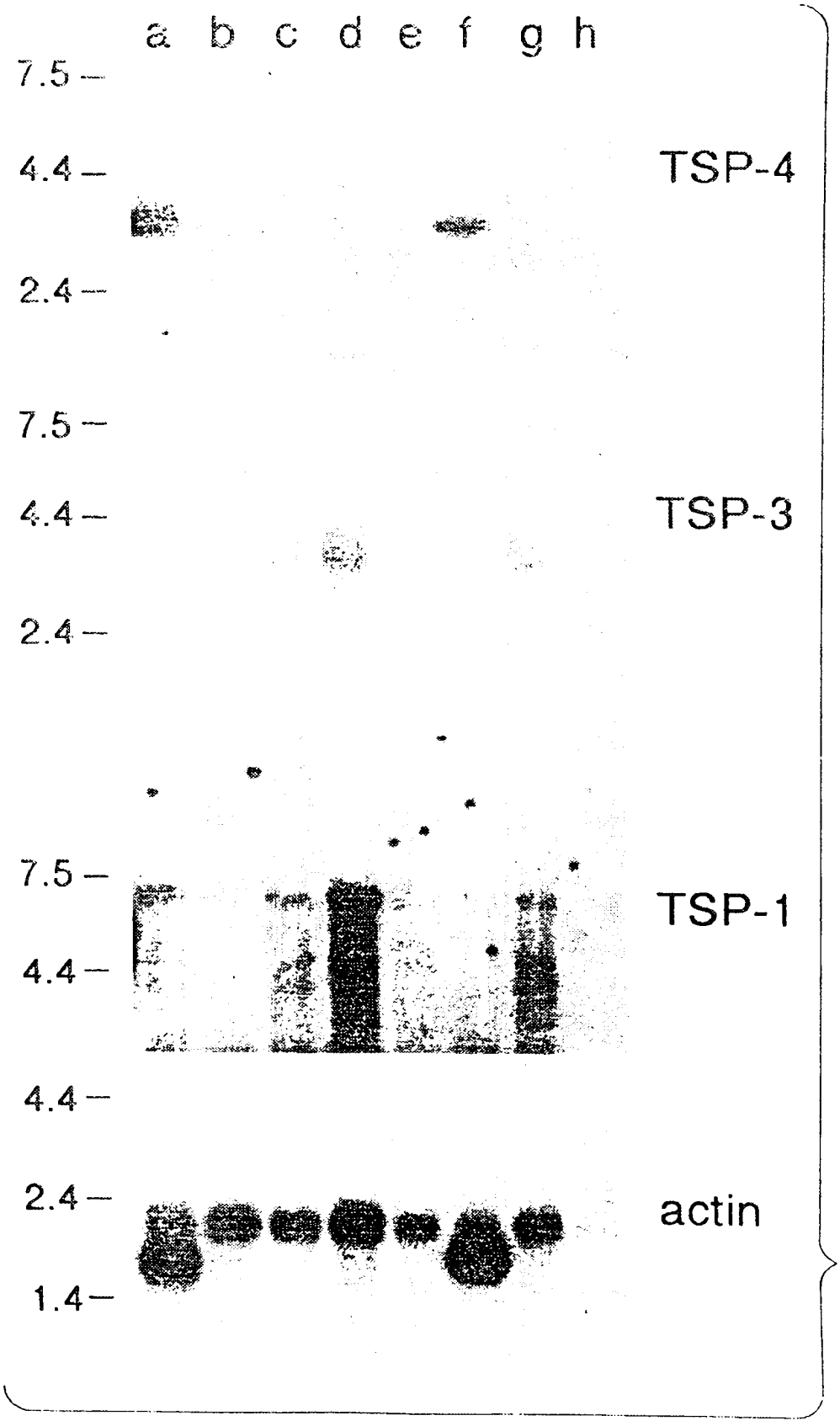


Fig. 5

INTERNATIONAL SEARCH REPORT

Internation pplication No
PCT/US 93/11725

A. CLASSIFICATION OF SUBJECT MATTER
IPC 5 C12N15/00 C12N15/12 C12N1/21 C07K13/00 C07K15/28
C12Q1/68 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 C12N C07K C12Q A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	J. BIOL. CHEM. vol. 267, no. 17 , 15 June 1992 , AM. SOC. MOL. BIOL., INC.,BALTIMORE, US; pages 12192 - 12196 H.L. VOS ET AL. 'Thrombospondin 3 (Thbs3) a new member of the thrombospondin gene family' cited in the application the whole document --- -/--	1-46

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 March 1994	20. 05. 94
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Authorized officer Hornig, H

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	J. BIOL. CHEM. vol. 266, no. 20 , 15 July 1991 , AM. SOC. MOL. BIOL., INC.,BALTIMORE, US; pages 12821 - 12824 P. BORNSTEIN ET AL. 'A second, expressed thrombospndin gene (Thbs2) exist in the mouse genome' cited in the application the whole document	1-46
A	J. BIOL. CHEM. vol. 265, no. 27 , 25 September 1990 , AM. SOC. MOL. BIOL., INC.,BALTIMORE, US; pages 16691 - 16698 P. BORNSTEIN ET AL. 'Characterization of the mouse thrombospondin gene and evaluation of the role of the first intron in human gene expression' cited in the application the whole document	1-46
P,X	J. CELL BIOLOGY vol. 120, no. 4 , February 1993 , ROCKEFELLER UNIV. PRESS, N.Y. , US; pages 1059 - 1067 J. LAWLER ET AL. 'Identification and characterization of thrombospondin-4, a new member of the thrombospondin gene family' see page 1064, right column, line 12 - page 1066, left column, line 25; figures 3,7	1-11, 13-18
P,X	CURRENT BIOLOGY vol. 3, no. 3 , March 1993 , LONDON,GB; pages 188 - 190 J.ADAMS AND J. LAWLER 'The thrombospondin family' the whole document	1-11, 13-18
P,X	J. MOL. EVOLUTION vol. 36, no. 6 , June 1993 , SPRINGER VERLAG, NEW YORK, US pages 509 - 516 J. LAWLER ET AL. 'The evolution of the thrombospondin gene family' the whole document	1-4, 15-18

2